A Feature-based Cost Estimation Model

by

Narges Sajadfar

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

ENGINEERING MANAGEMENT

Department of MECHANICAL ENGINEERING

University of Alberta

Abstract

To address the requirement of dynamic pricing and cost control in high-variation product manufacturing, nowadays many companies face the problem of generating quotes and order prices timely, accurately and consistently. This thesis reports a preliminary investigation on automatic cost estimation with a feature-based semantic model.

A generic semantic model for the purpose of automatic cost estimation is proposed, in which a new concept of *cost feature*, is suggested. A cost feature can be identified with data mining methods for different clients or products, and conceptually interfaced with product design and manufacturing features. Feature-based mapping models can be used to determine feature scope and cost level definition, including all the dependency relationships with other domain features. This model is expected to enable a visual, flexible and semantically consistent scheme to address effective and efficient product cost structures, frequent configuration variations and business changes.

Cost feature has been defined in this work as a unique class in the unified feature modeling system to address the characteristics of cost engineering entities, constraints and dependency relations. This research also describes the relationship between a cost feature and three engineering sub-models, i.e. machining model, design model and other auxiliary data model by associating tangible and intangible data. Further, semantic relations are investigated in early product design process for dynamic, accurate and

visible product cost estimation. This research also discusses cost engineering related functions, associated data structures, and techniques to be used.

Further, this research presents a Cost Estimation (CE) method that is tailored to apply feature-based engineering concept with data mining algorithms. The method proposed combines and leverages the unique strengths of linear regression and data mining approaches, and is based on a mechanism to discover cost features. The final estimation function takes the user's confidence levels up by phasing in the real time application of the method gradually and building up the data mining capability.

A case study is presented to demonstrate the hybrid method, and it compares the results of empirical cost prediction via data mining. The case study results indicate that the combined method is powerful and accurate for determining the costs of the illustrated welding features presented. With the results of the empirical prediction of five different data mining algorithms, the most accurate algorithm for welding operations is identified and presented in more detail.

Preface

This thesis is an original work by Narges Sajadfar. The semantic modeling approach that used in chapter 3 and 4 was published in two book chapters:

-Sajadfar, Narges, Yanan Xie, Hongyi Liu, and Y-S. Ma. "Introduction to Engineering Informatics." In Semantic Modeling and Interoperability in Product and Process Engineering, pp. 1-29. Springer London, 2013.

-Sajadfar, Narges, Yanan Xie, Hongyi Liu, and Y-S. Ma. "A Review of Data Representation of Product and Process Models." In Semantic Modeling and Interoperability in Product and Process Engineering, pp. 31-51. Springer London, 2013.

A version of chapter 3 has been published in Journal of Engineering and Technology. Ma, Y. S., Narges. Sajadfar, and Luis. Campos Triana. "A Feature-Based Semantic Model for Automatic Product Cost Estimation." IACSIT International Journal of Engineering and Technology 6, no. 2 (2014).

A version of chapter 4 has been published in International Journal of Mechanical Engineering and Mechatronics. Sajadfar, Narges, Luis Campos Triana, and Ma, Y. S.. "Interdisciplinary Semantic Interactions within a Unified Feature Model for Product Cost Estimation."

A version of chapters 5 and 6 has been submitted to Advanced Engineering Informatics. Sajadfar, Narges, and Yongsheng Ma. (Revision)

Technical case study and data collection were provided by McCoy Drilling and Completions (Farr) Company.

Acknowledgement

I would like to express my greatest thanks and appreciations to my supervisor, Prof. Yongsheng Ma; for his guidance, help, and support. I am very grateful for his invaluable contributions to my personal and professional development. I could not have asked for a better supervisor.

I express my deepest thanks to McCoy Drilling and Completions (Farr) Company, for giving me the opportunity to complete MITACS Accelerate Cluster internship program for two years. I would like to thank McCoy supervisors: Jose Pazuengo, Julio Barradas and Luis Campos for their tremendous support. I would like to extend my appreciation to my coworker at McCoy: Yanan, Tony, Henry, Parthiban, and Pedro. They created an enjoyable working environment.

I am so grateful to all my lab mates and friends in MECE 4-23 lab for creating a warm atmosphere: Fatemeh, Qingqing, Lei Li and Moin.

My special thanks go to my dear friend Afsaneh Esteki who helped me to learn and work with data mining software by sharing her knowledgeable advice.

I would also like to take this opportunity to thank colleagues and friends, Nahid, Fatemeh, Fahimeh, Neda, Leila, Parisa, Sara, Roya, and Negar for their support, care and unforgettable friendship.

And finally but importantly, I am also thankful to my parents and my sisters for all their love and unconditional supports throughout my life. Also, my sincere appreciation is given to my husband Mohammad Mahmoudi, for his encouragement and continues support.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

In today's competitive world and the on-line business era, effective and timely cost estimation has a critical role in manufactory's success. In terms of practical applications, industry needs to quickly estimate the price for any new product configuration in a customer order either interactively or automatically by on-line applications. High cost estimation accuracy at the very early stage and effective cost control during the operation phase are essential for the smooth business operation. Quick and accurate cost estimation brings lots of advantages over competitors. Cost estimation can also affect all aspects of production processes and management decisions, such as choosing between producing the product components in-house or outsourcing them.

The common challenge of most companies is to improve profitability. Historically, timely and accurate cost estimation has an important role in a manufacture's success that can improve the manufacturing performance and effectiveness of business [1]. Companies follow different strategies and practices that can increase the revenue and decrease expenses. Cost engineering research is imperative to guide companies by developing tools and models for effective and efficient cost estimation and cost control.

Cost estimation is an important process for management decision making. For example, the manufacturing manager needs to provide detailed and accurate cost estimation for making the strategic decisions, such as determining the bid price for product, make-or-buy decisions, marketing approach and setting the goals of profitability. To achieve all the benefits of accurate cost estimation, the manager needs to formulate the best suitable cost estimation method that best matches the specific products' structures and individual business nature.

In addition, cost estimation can be carried out in each stage of a product life cycle, such as conceptualization, concept design, detailed design, prototyping and manufacture. Many researches recognize that up to 80% of the product cost is determined at the conceptual and design phases [2]. Therefore, accurate cost estimation at early stages with different levels of defined details has an important role in manufacture success.

1.2 Problem statement

In manufacturing, accurate cost estimation and tracking cost estimation are well-known problems, and a definite solution to these is nonexistent. The traditional cost estimation approaches focused on parametric, analogy and bottom-up techniques and they are still being used in many companies of different industries [3]. Most financial managers rely on empirical cost estimation as a decision tool. However, in today's competitive world, with a large variety of business models, products and purposes, cost estimation should be done in a comprehensive system that can include all the details, and calculate the cost automatically and accurately.

In the past two decades, Enterprise Resource Planning (ERP) systems have been essential contributors to the success of costing advancement in manufacturing companies. The development and implementation of ERP systems helps industries to gather and manage all their information into a common computer-based information system [4]. By utilizing integrated ERP systems in manufacturing, it seems true that most of the cost calculation issues were addressed [5]. The ERP cost calculation can be typically for a specific segment in a production cycle, for specific machining and manufacturing process, or for a specific part and component. ERP cost calculations are based on parametric cost estimation that requires all data for the details of parts or processes, but such complete data is usually not available, particularly in the dynamic design and early sales states where prototypes are being built and configurations continuously modified.

In addition, manual variations of processes, such as welding, are not well defined and documented on ERP systems; and companies with small batch size and high variation production have difficulties to model product costs accurately.

This research addresses cost data availability and verification issues by proposing a new cost estimation modeling approach based on a semantic feature-based cost estimation engine, together with the use of a cost feature library and system integration. However, such implemented methods are currently limited to simple cost estimation, and depend on approximations and assumptions. Even the most complete ERP system has numerous limitations regarding the cost estimation. For instances:

- Specific organization or manufacturer cost scenarios are not considered in ERP costing data setup [6]
- All cost-details data have to be provided for accurate cost calculation which requires a lot of time, and the information used may not be accurately assessed
- Missing data has a huge negative effect on cost estimation accuracy and calculation
- The existing ERP systems could not calculate the cost base on products configuration and machining features
- There is not any reusable process for gathering information and cost calculation
- The historical data is often not considered in ERP cost calculation
- The particular company costing strategy is not effectively reflected in ERP cost implementation
- The calculation of a new product cost is not accurate due to lack of information

1.3 Research challenge

Manufacture cost estimation is a cluster of challenging tasks done in three levels: conceptual design, detail design and manufacture. Although cost estimation is clearly related to the profitability, the systematic pricing was not adopted in most companies due to the complex associativity among the technical and business considerations. A systematic semantic model is required in order to tackle the relations and constraints among product elements, processes, resources and actors.

Many manufacturers used to do their cost estimation at the manufacturing process level; this approach represents the traditional way, and is also known as engineering bottom-up method [7]. In the manufacturing process level, all the expenses can be quantified already. For example, Donald R. Woods mentioned more than 20 factors in his book, such as: cost of fabricated equipment, cost of material, cost of labor, cost of control, cost of land, cost of storage, legal fees [8], etc. Although this traditional approach is still being used, however its drawback is pretty clear, that is when the product gets complex, the accuracy of cost estimation will decrease significantly. Also there are lots of uncertain factors behind all the cost elements, such as the changes of costs related to the changes of materials, processes and equipment items. Such uncertain dynamic changes can increase the risk of broad range of cost estimation, such as the risk of misleading loss-indication in a potentially profitable deal [9], or the opposite.

Cost estimation is a dynamic procedure that must be done periodically during each stage of a product lifecycle; and it has been the trend to compete on price at product conceptual specification and the early design level. So instant and yet more accurate cost estimation is demanded at the early stage of product development. However due to the lack of accurate information in the early stage of a product, or a business order, the cost estimation is currently carried out with experience and subjects to significant modifications at the manufacture level.

The purpose of this research is to introduce a new semantic model that can systematically associate design features and manufacturing features with a cost estimation mechanism, i.e. the proposed cost features. They can serve as the active information agents and work out cost elements intelligently in the product specification stage with the inference capability from the historical and expected cost models of associated features at the design and manufacture levels.

1.4 Research goals

The top objective of this research is to improve the accuracy of production and machining cost estimation by defining a set of cost features and developing a prototype of a cost estimation engine. This thesis also reports the implementation of new feature-based cost estimation method by using the cost feature concept and a hybrid data mining and empirical algorithm. With the proposed approach, cost estimation can be partially automated with a gradual adoption process with the user-controlled. The feature-based cost estimation method involves the following steps:

- Identifying ERP cost estimation limitations
- Develop a method calculating manufacture costs by identifying the processes, involved machines and tools for each product component and manufacturing processes
- As a test case defining cost estimation scenarios with newly defined welding features
- Gathering all the cost-related information and deposit the data into a database
- Developing a computer-based cost simulation tool and using it to estimate product cost scenarios

1.5 Research scope

This research, presents a novel feature-based cost estimation method by combining two advanced cost estimation approaches, empirical and Data Mining (DM) by using cost feature concept. A prototype software toolkit is also implemented. Cost feature is defined as a set of characteristic attributes of an object that can be associated and constrained to represent the patterns of semantics for the purpose of cost engineering. The automation of cost estimation based on cost features includes three steps:

- Gathering historical data:

  Our selected company's ERP has the information about sales, inventory, purchasing, and manufacturing of products. Based on project needs, we can transfer information from the company's ERP into our new database, where we have more control on the existing data and we can do the necessary analyses.

- Abstraction of cost feature and cost feature data extraction:

  Each product has different characteristic engineering attributes such as: design features, functions, suppliers, tooling, machine profiles, machining features, etc. and each of this attributes have their own influence on cost. Ideally, the cost estimation feature could make predictions based on product functions (f) and their weights (W) for each manufactured product. For example if a product includes components A, B, and C. The cost feature for this product is:

  $$C_{total} = W_A * f(A) + W_B * f(B) + W_C * f(C) \qquad (1\text{-}1)$$

  Abstraction of feature related to costs should be generically workable for all different parts, modules, sub-assemblies and products that is describe in chapter 5 and 6 with more details.

  The database that we have developed for this project contains information for Bill of Materials (BOM) and Method of Manufactures (MOM) of 500 different manufactured products from the company partner. Additionally, the cost engine has to recognize and extract cost feature data in order to obtain the cost estimation.

- Using different algorithms and methods to find the most accurate cost estimation
  Based on data mining (DM) techniques, the cost engine has to select one effective cost estimation algorithm out of several available options to gain a trustable cost estimation result with measurable confidence level. WEKA is a machine learning software tool that provides a comprehensive collection of algorithms for automatic classification, regression, clustering, and feature selection [10]. WEKA has been used in this research as a tool in order to support the cost feature identification and extraction, and application. The DM results provide a guide for selecting the most accurate algorithm based on the available data sets.

To complete the evaluation cycle for the proposed approach, the data set collected from ERP historical records is divided into three subsets, stored in separated databases: training data, testing data, and application data. WEKA uses the training data to generate specific algorithms. Then WEKA uses this algorithm on the testing data to estimate the costs. With the testing database, the actual costs are also known, and then the reported costs can be compared with the results from the proposed cost estimation algorithm. If the accuracy of cost estimation is acceptable, then WEKA algorithm is used to estimate the costs of the products in the application database, then the final effectiveness is measured by a confidence level for the feature application of the cost estimation engine.

1.6 Research methodology

Feature technology has been successfully applied in engineering computer tools in the past decades because of its capability to resemble engineer's semantic patterns related to their design and manufacturing threads. Feature is a kind of well-defined and dynamic data structures that can describe more information related to all the aspects of the product that is not reality described in traditional analytical models or semantic attributes [11]. Many advanced engineering design features and machining features have been successfully defined and applied. Features can describe the material characteristics, machining methods, geometric shapes, etc.

The feature-based cost estimation approach entails to identify cost related features that can be defined as associated to certain functions and manufacturing processes. For example, a part's cost can be linked to the design features for the estimation of material cost, power, optional add-ons, and other quantifiable impacts. Cost can also be linked to manufacture features to estimate the machine utilization and process time cost [12]. However, so far in practice, identifying, defining and the tracking the dependencies and impacts of changes related to *design and machining features* at each level of cost estimation are very complicated procedures and almost impossible for a complex product.

1.7 Thesis organization

Relevant methodologies and techniques in the areas of cost estimation and feature-based modeling which are previously investigated by other researchers are first reviewed in Chapter 2.

Chapter 3 discusses the concept of proposed approaches by introducing research challenges and methodology. The goal of this chapter is to investigate a concept of new manufacturing cost calculation model coherently throughout the lifecycle of a product series, especially emphasizing at the conceptual design stage with the input of expected manufacture process information. The proposed model integrates three functional sub-models: feature-based costing, data mining, and semantic reasoning.

Chapter 4 is devoted to a more comprehensive description of the proposed system:

- A data structure that uses the unified feature concept to support data associations and sharing across the design model, machining model, and auxiliary data model.
- Inherent relations among cost feature and other features, such as design features and machining features.

The purpose of this chapter is describing a new cost estimation functional module and its semantic interactions within the above-mentioned three sub-models and illustrating the effective definition and use of the cost feature objects which can be directly usable in computer functions for the cost estimation at different engineering phases or levels, such as the design and manufacture stages.

Chapter 5 focuses on the discussion of tools and procedures used to carry out the hybrid cost estimation method based on feature-oriented data mining:

- Definition of cost feature and its requirement
- Empirical and data mining cost estimation processes
- Implementation of data mining algorithms in the proposed cost engine

Research result and case study are discussed in Chapter 6 followed by conclusions and future work recommendations.

Chapter 2: Literature review

2.1 The need for cost engineering

The American Association of Cost Engineers (AACE) describes cost engineering as "that area of engineering practice where engineering judgment and experience are utilized in the application of scientific principles and techniques to the problem of cost estimation, cost control, and profitability" [13]. Cost engineering tools help companies with decision-making, cost management and budgeting control [14]. Cost engineering includes cost control, cost forecasting, risk analysis, investment evaluation and profitability analysis. However, cost engineering focuses more on cost estimation and cost control.

In addition, cost engineering during the early stage of product lifecycle has an important role in product development. Many researchers agree that between 60% to 80% of a product cost comes from the concept phase [2, 15, 16]. Figure 2.1 shows a typical cost commitment curve during the concept and production phases. In average, decisions made during concept design phase contribute up to 70% on the total cost of a finalized product. However, the actual cost incurred in the conceptual design level is less than 20% when customized products are developed, specification of configurations affect the total cost significantly [17].

Also, many decisions need to be taken during the further stages of the product development cycle such as: make-or-buy, manufacturing process plan, recycling of the product, etc. The decisions are based on different criteria and cost estimation has an important role to influence the decisions. However, limited amount of available data during the early stage, uncertainly variables and unknown conditions make the cost estimation accuracy questionable.

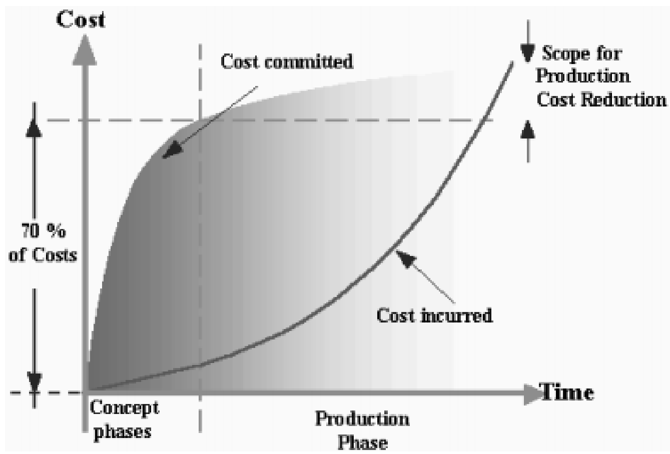


Figure 2.1 Cost commitment curve [17]

2.2 Cost types

The definition of manufacture costs is the sum of all costs of the resources (labor, equipment, tools, materials, energy and etc.) that was used in the process of making a product [18].

All the manufacturing costs can be divided into two main categories: fixed costs and variable costs. Fixed costs are those cost that do not change based on a quantity and it's independent of output. On the other hand, variable costs are based on the products quantity and it's growing by increasing the output [19]. In addition, cost categories can be broken into direct and indirect costs. Direct costs are directly and consistently connected to the production of the product such as machining operations, service costs, software and tool. Indirect cost cannot be recognized or pinpointed for specific products or parts such as building rent, utilities, administration, maintenance, and staff salaries [20]. Figure 2.2 illustrates the relation among fixed and variable cost and quantity of products. Fixed cost has a flat line while variable cost is increasing by number of products.

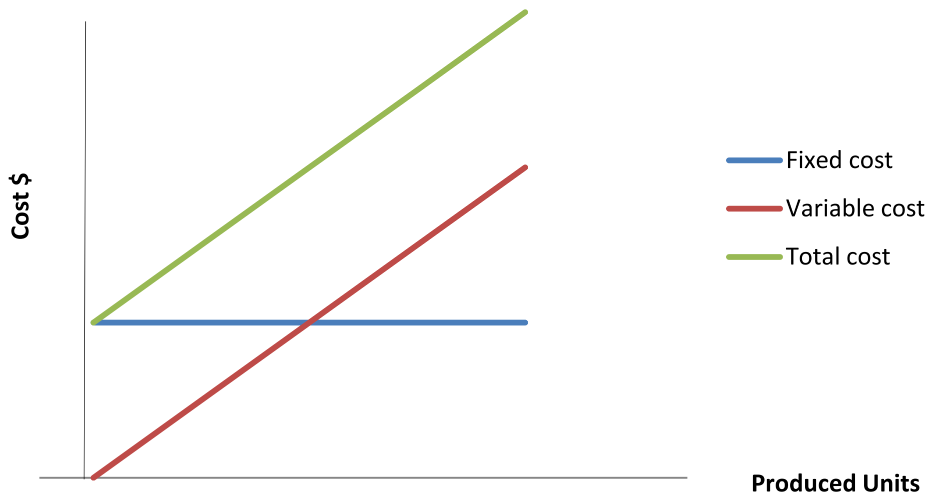


Figure 2.2 Fixed and variable cost

Different categorization methods for cost items can be formal in [21]. Table 2.1 shows an example of total product cost categories which introduced by Asiedu, 1998 [22]. At the first level, total cost is divided to four main categories: research & development cost, production and construction cost, operations and maintenance cost and retirement and disposal cost. Next level describes more details of each category. For instance:

- Product software belongs to research and development cost which is direct and fixed cost.
- Quality control belongs to production and construction cost that is indirect and variable cost.

Table 2.1 Total product costs category [22]

Total products cost			
Research & development cost	Production and construction cost	Operations and maintenance cost	Retirement and disposal cost
- Product management - Product planning - Product research - Design documentation - Product software - Product test and evaluation	- Manufacturing/ Construction management - Industrial engineering and operations analysis - Manufacturing - Construction - Quality control - Initial logistic support	- Operations/ maintenance management - Product operation - Product distribution - Product maintenance - Inventory - Operator and maintenance training - Technical data - Product modification	- Disposal of non-repairable - Product retirement - Documentation

2.3 Classification of product cost estimation techniques

There are various techniques and methodologies for product cost estimation with their own advantages and disadvantages. This chapter contains a review of the product cost estimation literature, and focuses on the techniques and software tools used. We can classify cost estimation techniques (CETs) into different ways, such as: traditional and modern ones in the sense of adoption time; qualitative and quantitative from the analysis methods; top-down and bottom-up from reasoning approaches; and parametric and judgmental cost estimation from data input and propagation angle.

Traditional cost estimation approach includes bottom-up engineering, analogy, and parametric methods. They are fundamentally based on historical data and experience knowledge. The advanced cost estimation approaches include expert judgment, feature-base evaluation, fuzzy logic, neural network and data mining (DM) methods; they are more suitable for complex products and dynamic cost estimation [23].

Regarding the accuracy of cost estimation adopted from [12], all the techniques can be divided into two groups as figure 2.3 shows: qualitative and quantitative ones. Qualitative cost estimation is used when it is difficult to accurately describe the cost based on the product data, and it requires heuristic methods [24]. Usually, this technique requires judgment of an individual with expertise to identify the similarity between the studied product and previous products in order to relate their costs [12]. Quantitative cost estimation is based on product details and processes data. Historical data and statistical models have a main role in quantitative cost accuracy [25]. In the next level, qualitative technique includes intuitive and analogical methods; and quantitative techniques include parametric and analytical methods [12].

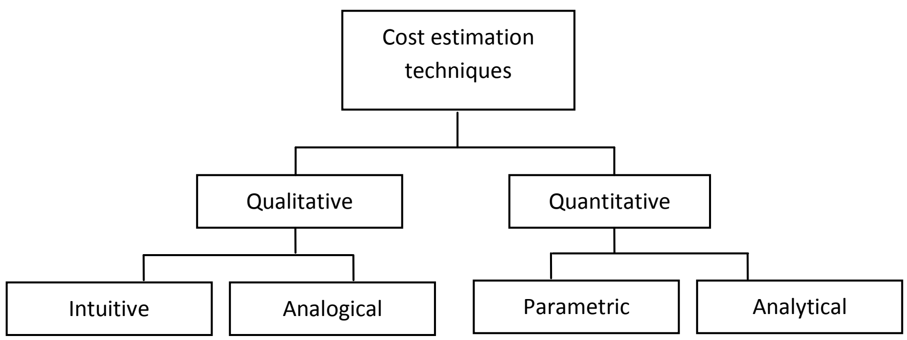


Figure 2.3 Initial classification of product cost estimation [12]

2.3.1 Qualitative techniques

Qualitative CE uses a heuristic method in the case of difficulty in finding the accurate cost based on parametric product data [24]. Intuitive cost estimation is the first category from qualitative CE which is a fast CET that use experienced individual for cost estimation. However, each individual (expert) is likely to include his personal input and thus is likely that two expert individuals will come up with different cost estimations;

which will certainly decrease cost estimations accuracy. To solve this issue the researcher has defined different systems and methodologies to create a more accurate framework for this CET [26]. Intuitive techniques can be further largely subdivided into two approaches: case-based reasoning and decision support system-based.

- Case-based reasoning (CBR) uses an expert person to find a past design similar to the new design and associate their costs. The past design will be adjusted to become similar to the new design, and then the cost of each change will be calculated [27]. A smarter way in this approach is using an information processing technique to extract geometrical features from the product Computer-aided design (CAD)-model. Then the similar cases are identified and the costs suggested by a system for cost calculation [28]. Usually this technique is used through conceptual phase or design phase to decrease the total cost of the product.
- Decision Support System (DSS) is a management tool, which can be defined as a theoretical system to describe the processes of decision-making or the computer-based system that is used to represent the information of a decision-making process [29]. DSS can be used for cost estimation purpose and implemented based on three techniques: rule-based system, expert system and fuzzy logic system. Each of these techniques provides a framework for decision-making process in cost estimation and needs to have manufacturing process knowledge and practical experience to make it practically usable [30].
    - Rule-based system: This is the most common technique for DSS, which is an Artificial Intelligence (AI) application. This technique needs to identify a pattern and structure from data to identify applicable situations, and then it has to define an appreciate action for each situation. This AI technique was introduced in 1960s, but this method was not used for cost estimation until the end of 1980s [31]. Kingsman Brian et al, [30] identified around 200 heuristic cost estimation rules for make-to-order system. Gayretli, A. et al, [32] introduced a rule-based system for feature-based cost estimation. They extracted the form feature and process feature from CAD model, and then

the available processes are suggested by rule-based system based on feature attributes process capability and production rules. In the same time, the rule-base system will calculate the cost based on the programmed functions.

- Expert system is the first commercial artificial intelligence program that simulates the human judgment by using expert knowledge. The purpose of using expert system is to help general managers to make better or faster decisions, or it can be used partially instead of a professional for basic decision making in a specific area. Expert system contains two main parts: interface engine and knowledge base module. Knowledge can be store based on different facts and rules into knowledge-based module and interface engine suggests the solution based on the pre-defined rules stored in the system [33].

Analogy cost estimation is a qualitative technique that compares the new product with the past product based on functions and geometrical similarity [25]. The degree of similarity has a direct effect on accuracy of analogy cost estimation. Usually, past products require changes to arrive the new product information which needs time and has to be done by expert professionals. In addition, there is not a unified framework for analogy cost estimation and each expert person can do it in different ways. Thus, their results are not comparable. Analogical techniques are further subdivided into Regression analysis models and Back propagation neural network (BPNN) models.

2.3.2 Quantitative techniques

The quantitative CE approach is focused on analytical cost calculation functions and the exact parameters summarized from the records of business transactions for costing purpose [12]. Quantitative techniques have two main categories: parametric and analytical techniques.

Parametric cost estimation (PCE) is a principle method for cost estimation in early stage of product life cycle. Usually there is not enough information about the product in conceptual or design stage. PCE can provide a statistical relationship between variables based on previous product. As a result, the build statistical formula can be used for new product cost estimation [24]. In general, PCE tries to define the relation among product characteristic and cost information. Then, it uses the extracted methods for top-down cost calculation. Thus, it can be described as a parametric model that contains different formulae [25]. Many techniques and methods can be used to extract the statistical relationships. They can be categorized in three main methods: Statistical methodology, linear regression, approximate tool paths and process parameters. Linear regression is the most common technique using statistical methodology to measures the dependency of selected variables in relation to other independent variables [34].

Analytical techniques are focusing on detailed design. The purpose of analytical techniques divide the product based on their activities, operations and properties to create more details and information regarding to the specific parts or products [35]. Feature-based cost estimation and activity-based cost estimation are two common techniques in analytical techniques categories.

- Activity based:

Activity based costing (ABC) is a costing approach that outputs the cost of products based on the activity being performed on the product. ABC has two main central parts: activities and cost drivers. The cost of each activity has to be calculated first, and then it will be assigned to the product. The total cost will be calculated by adding each activity assign cost [36]. This type of CE is very useful for manufacture cost estimation which can calculate the cost based on the manufacturing activities. ABC is proposed by reference [37]. Through investigation on traditional cost estimation, they suggested to use ABC for calculating overhead cost instead of using fixed overhead cost.

- Feature-based:

Feature-based CE can be very much detailed and accurate for complex products; this method basically decomposes product manufacturing processes into cost factors based on their operation methods, features, surface quality requirement, key process dimension and geometrical tolerances, and supporting activities,. In order to manage the tedious cost elements more efficiently while keeping the accuracy of CE, the correlated cost patterns can be identified and categorized with product configuration variations and predefined procedures. In other words, analytical examination can be extended to a feature-based approach.

2.4 Review on feature-based cost estimation

In this research, a feature-based CE approach is proposed as an analytical CE technique to identify cost-related features that are associated with certain function specification and manufacturing processes [12].

Feature recognition goes back to Kyprianou's PhD work at Cambridge in the late 1970s. In 1980s, Shah et al. [38] introduced the geometric feature reorganization from CAD model, they mapped feature concept with engineering semantics and software engineering object-oriented approach conceptually. Later, Emmerik V et al. [39] introduced a set of graphical user interfaces for feature-based modeling. The feature-based engineering informatics approach has been gaining more and more popularity into computer-aided software tools, such as in CAD and CAM domains. However, this model has not been generally used for cost-determining purposes.

Leo Wierda [40, 41] is one of the first researchers who used the concept of feature in qualitative and quantitative CE. In 1990, he introduced a new tool that would enable the user to effectively control costs. He defined the "life cycle cost" by three phases in the model, including conceptual design, materialization, and detailed design. He also defined two types of cost information: design rules and manufacturability information. Design rules refer to the historical design information used to describe a framework for a current design. Manufacturability information refers to the current design information and geometry information of a product; this information is used to extract the product cost data. In his cost model, the target cost is considered in each phase. As a result, the design is based on cost, and the cost is based on information derived from the feature reorganization program [40]. In 1991, Leo Wierda expanded "life cycle cost" to include feature-based cost modeling. He used the feature-based model to fill in the gap between the design and the process plan. In addition, he described the advantages of calculating the costs per features [41].

By the end of the 1990s, more researchers focused on feature-based CE, and more techniques were introduced for using the feature-based cost estimation. Zhang et al. [42] launched the feature-based CE by using back-propagation neural networks, especially for packaging products. They introduced the definitions and the quantification of the cost-related features, and then they used the back-propagation neural networks to figure out the relationship between the cost-related features of product packaging and product cost. Ou-Yang and Lin [43] suggested new tools for feature-based CE in the early stages of the product lifecycle. They extracted factors that might affect production costs in an attempt to reduce the costs in the design phase. At the end, their method can calculate the cost based on products' shapes and the accuracy of the extracted factors.

Leibl [44] designed a program-based system for feature-based CE. The program integrates the CAD system to extract the geometrical and non-geometrical information from the CAD model; next, based on the degree of details that were provided for the program, the product's cost is calculated. This system has to use other design document to make a better cost calculation, as it cannot do a CE just based on the CAD model. Brimson [45] used the feature-based CE as an alternative activity-based costing (ABC). In this approach, the product's features have to be identified first; afterward, the working step or the requirement activity of each feature can be recognized. Finally, the cost of each activity has to be calculated.

In the last two decades, many researchers have investigated into this area to improve the efficiency and accuracy of feature-based CEs based on specific manufacturing operations. F. Masmoudi et al. [46, 47] used a feature-based concept to do analytical and parametric welding CEs with computer-aided tools. They described a CE model that uses preparation features and welding features to approximate an effective welding time and, thus, estimate the cost. They used the welding feature based on the forward reasoning calculation with multi-parameter inputs, and they captured more details than they would have taken with the traditional calculative approach. The limitation of this approach is that to make a parametric CE, all of the feature data is required, which is not easily available in practice.

In this thesis, a different approach is proposed using a hybrid method; in addition to taking a similar parametric and forward-reasoning, CE route, a backward data mining approach is also adopted as an alternative source of cost evaluation. Further, these two routes are merged by a weighted combination for the ease of application and better accuracy of prediction. The advantage of taking this hybrid approach is that historical cost data and cost features are used and matched with data mining techniques, and the backward approach could predict costs, even if all of the data has not been provided.

It is important to know that to enable accurate prediction of feature-based manufacture cost, costs incurred should be related to the current plant capacity status and the forecast, which can only be made available from the shop floor. Few CE systems consider this issue [48]. In recent years, several CE models have been introduced to consider capacity limitations [48, 49]. In addition, Wei and Ma [50] introduced the capacity feature concept to fill in the gap between customer orders and resource management, which is considered an important real-time input resource for cost estimation.

Besides selecting the appropriate CE approach, the cost estimation software is required to improve the accuracy and consistency of the proposed approach. Currently, a large portion of the software for cost estimation (SCE) are using data mining algorithms or artificial intelligence (AI) techniques [51].

T.L. Karen et al. [52] reported about the effects of data mining techniques on cost estimation; they described that data mining techniques are not always implemented better than traditional cost estimation. However, a combination of data mining techniques can provide accurate SCE for specific products or processes [52]. The goal of this research is to introduce the combination of data mining algorithms that is suitable for feature-based welding cost estimation, which can be used as a tool for ERP systems.

2.5 Review on software tools for cost estimation

Software cost estimation is an important domain in software engineering practice. Different methods and techniques are using in software cost estimation. Jorgensen M et al., [53] classified these techniques as a following order:

1. Regression: regression analysis is a statistical analysis that can shows the relation among the dependent and non-dependent variables. Regression analysis can be used in purpose of minimizing the sum of relative errors and illustrating the non-negative coefficient [54]. Some cost estimation software's are using regression analysis for cost estimation. For example Constructive Cost Model-COCOMO is one of the algorithmic estimation models that are using basic regression formula [51].
2. Parametric: use the relationship between variables to calculate the cost

   Analogy: Analogy cost estimation is a common method that uses similar and completed product information to simulate the current product cost. The cost related information of previous product is using as a basis for new product.

Also this method is using for validating the results of the other cost estimation methods [55].

3. Expert judgment (EJ): based on the results of referring to the historical cost data the software can do the cost estimation.
4. Work break-down structure (WBS): identify each process/activity that generates cost
5. Function point (FP): the cost of single unit is calculated from past projects. It is based on the number of functions the software has to fulfil.
6. Classification and regression trees (CART): CART is one of the most important machine learning tools for constructing prediction models. Classification tree is using specific rules for splitting data at a node based to divide target variables into meaningful subsets and then regression tree is using to predict the value in each terminal node [56].

Chapter 3: A feature-based semantic model for hybrid production cost estimation*

3.1 Cost estimation model

Historically, attaining accurate cost estimation is a critical factor in business success, because cost engineering has a great effect on the manufacture profitability and management. Timely and accurate cost information helps business to make good decision and reduces business risks. Also, accurate cost estimation can create a competitive advantage. A good cost estimation method has a direct effect on sale price, amount of sales, market share, and enterprise profits [12]. The cost management function can be developed within the common Enterprise Resource Planning (ERP) systems. Those successful companies that can offer the best price within the short time on market will increase their market share. Lots of techniques and methods are introduced by researchers for product cost estimation in recent decades. There have been number of researches about cost estimation modelling and most of them focused on parametric and historical data [57].

Feature-based cost estimation is an analytical cost estimation technique which is more focused on quantitative approach rather than qualitative approach [12]. The main goal of feature-based cost estimation is to calculate the product total cost based on identified cost-related features. The most quoted limitation of feature-based cost modelling is the difficulty to maintain the semantics cost related of features during the product lifecycle.

---

* A version of this chapter has been published in Journal of Engineering and Technology. Ma, Yongsheng, N. Sajadfar, L. Campos Triana "A feature-based semantic model for automatic product cost estimation".

To avoid this limitation, the complex cost-related features, and the relations among other types of features have to be managed consistently.

In this chapter, cost feature has been defined [11] as a unique category of features in a unified feature scheme [58] that can cover all types of features which are related to cost estimation. As a result of associative feature modelling [59] which could support the cost feature concept, then using just one kind of features generically so that it can be well defined class to be semantically maintained during the product life cycle is a feasible and effective approach. In such a way, the properties and behaviours of other extended features defined on top of the generic feature model will be updated automatically. Figure 3.1 shows three sub-models, i.e. design model, machining model, and all other auxiliary data including tangible data, which refers to the set of data that has been computerized and can be extracted automatically by certain program interfaces, and intangible data which requires user's input interactively on spot based on human experience and knowledge. These three sub-models have conceptual semantic relations with each other's.

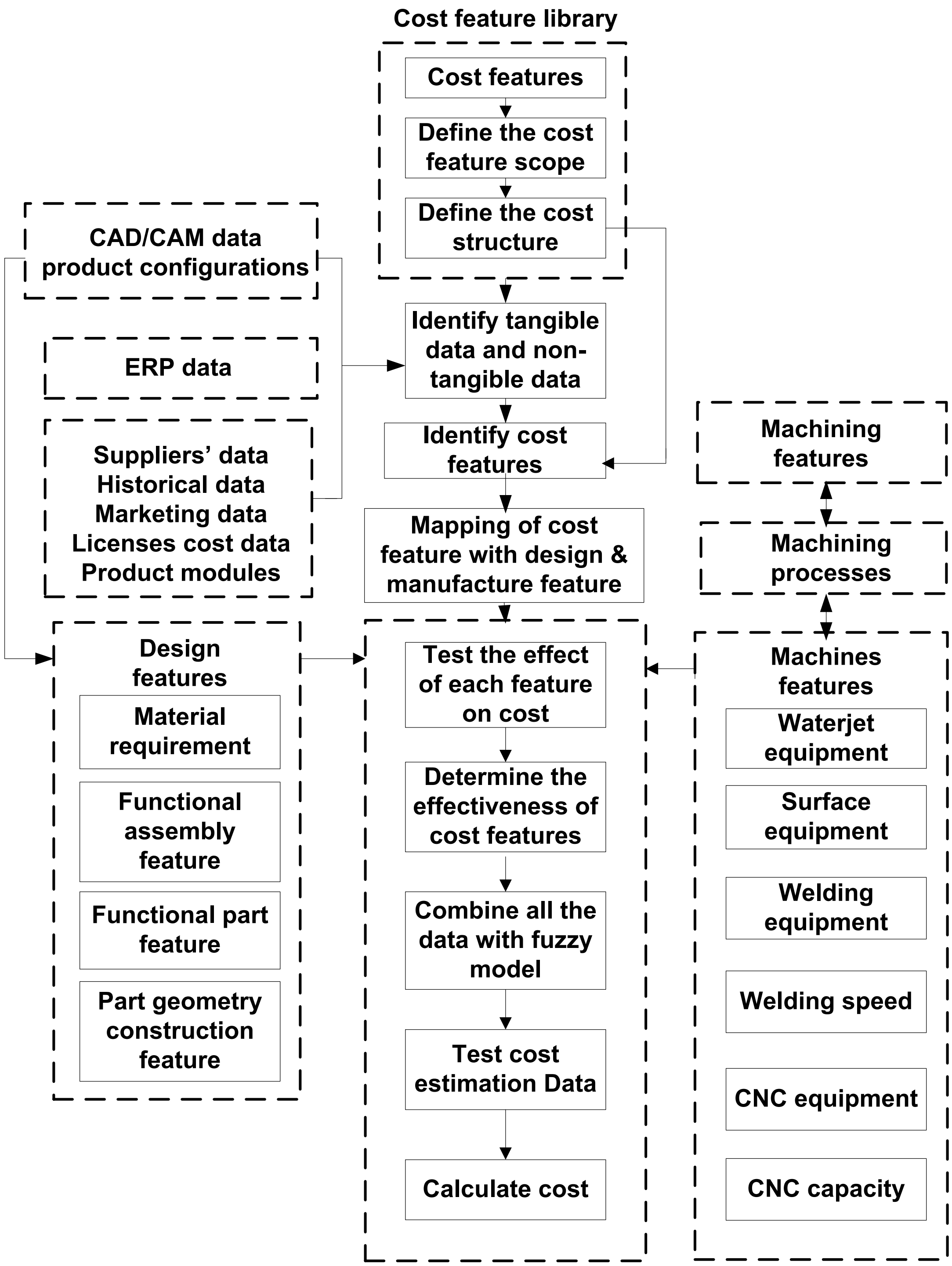


Figure 3.1 Cost feature related elementary semantics with some non-exhaustive relations

3.2 Cost feature and cost feature association

The candidate contemplates to define a new type of features, that is *Cost Feature*, a feature object class according to object-oriented software engineering paradigm, aiming to cluster, encapsulate, update and manage those cost-related dependencies, rules, constraints and referenced features defined from other domains, such as those related to design functions and machining processes. The *Cost Feature Association* method is introduced in this work to map the cost features with other types of features in cost estimation via a set of associative and semantic trees.

According to the figure 3.2, different types of information is required For identifying cost feature, such as the historical data, uncertainty factors, design and manufacture conditions, etc. Knowledge-based techniques like *Data Mining* (DM) can recognize all the product information from historical data, such as their attributes and the correlations and relationship patterns among their attributes. The next phase is cost feature establishment with attribute selection that is one of the most important steps. Given the fact that not all the information of the product is useful for cost estimation, most effective attributes have to be selected for this purpose [60]. Such selected attributes are then associated by using a decision tree. In the next step, the rules for data analyses and weight of importance are set. Finally, DM can present the data pattern that can predict the future information from historical data. Some useful algorithms are defined, prototyped and built into the methods of cost feature class as well as the derivatives of their object instances.

Also, other supporting functions of the cost feature are developed to enable the intelligent automation procedure and to persistently store information patterns as well as the attributes that was discovered by data mining, such as manufactory specification and inventory information.

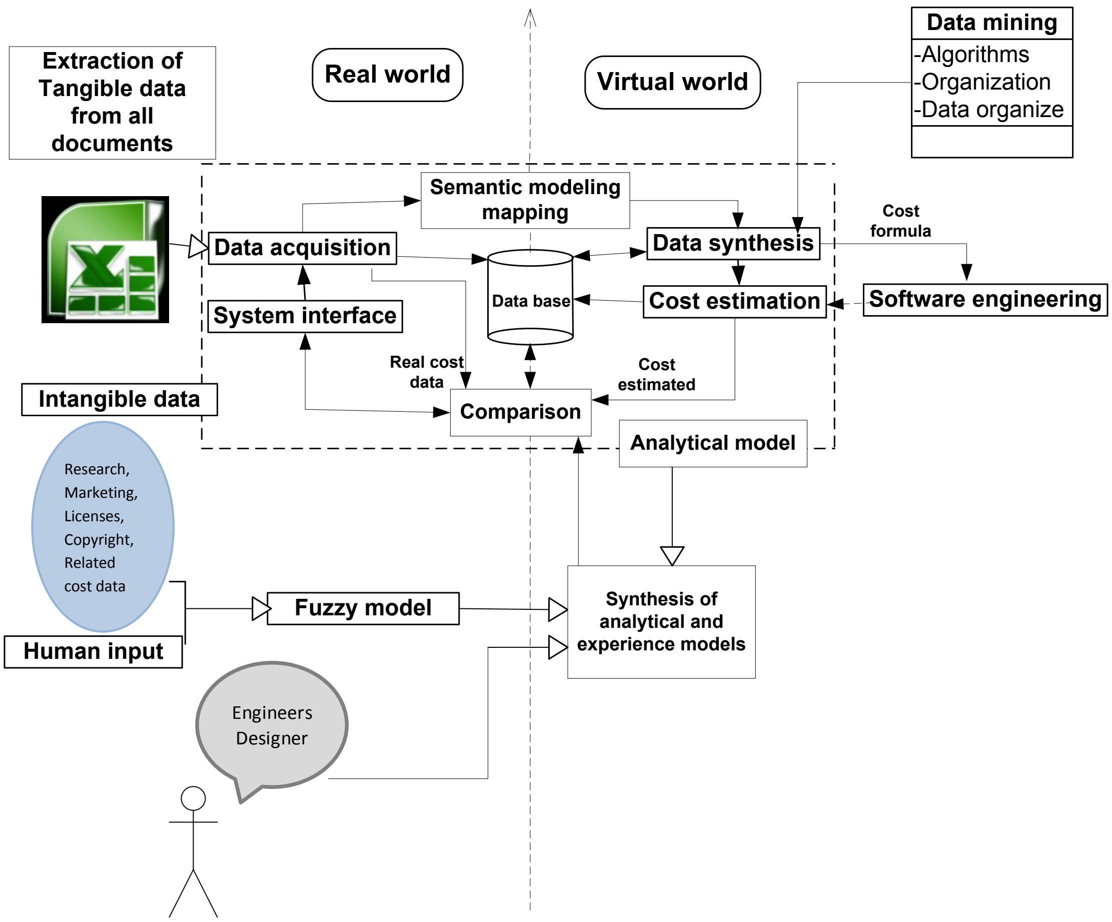


Figure 3.2 Concept of semantic processing model for the proposed cost estimation

3.3 Semantic modeling

To explore the scenarios of applying cost features, semantic modeling is necessary. Semantic modeling is a knowledge representation approach that expresses real world objects in the form of predefined terms and entities that are interpretable by a computer program such that the real world entities and their dynamic behavior can be consistently described, modified and persistently reused within the prescribed scope of application. Semantic modeling is an extension of high level object-oriented and modular software development method that is useful for consistent software engineering modeling. It is one kind of knowledge modeling that describes the meaning of data from the point view of the computer entities with interactions and user interfaces.

The objectives of semantic modeling are abstracting the knowledge and understanding the meaning of data and information. The other aim of semantic modeling is implementing highly cohesive software engineering model [61].

Semantic modeling is useful for mapping the entities from real to virtual world; modeling behaviors from complex life cycles of business, describing the dynamics of activities and illustrating the processes so that well-structured software elements, e.g. foundation classes, intermediate modules and functioning applications are modeled and implemented by object-oriented programming. Figure 3.2 shows the overall concept semantic modeling for cost estimation.

A disciplined semantic modeling approach for system implementation needs to follow several rules:

1. All the steps of semantic modeling have to be defined clearly.

2. The abstract model should create relations for mapping the information between the real and the virtual worlds.

3. The modeling objects have to be generically defined and directly mapped into class diagram.

4. A semantic model has to specify the comprehensive relationships and interactions among the modeling objects to describe the logics [11].

Three kinds of data are analyzed, classified, and related with some predefined relations. The first kind is referred to as tangible data that are explicit attributes related to cost objects describing properties such as physical, natural and economic data. On the other hand, another kind, intangible data are those attributes that are not obviously or directly related to cost elements but implicitly associated that the relationships are to be discovered and defined dynamically like the data describing context, background, social status, organizational attributes and cultural data, for example: copyright and licenses have indirect cost for organization. The third kind of data is interactive and subjective that requires engineering information and knowledge from marketing managers, designers, process engineers, and purchasers; this kind of data fills the gap between the tangible and intangible data types [62].

For doing reasonable and accurate cost estimation, all these three kinds of data are needed. Tangible data can be materialized in manufacturing view with CAD/CAM data models, historical data, ERP data, geometry data, manufacturing capacity, assembly and completion data. Intangible data includes brand value, goodwill, training, skill requirement, marketing positioning, software licenses, IP copyrights, research and development cost, etc.

In the next step, the conceptual semantic relations illustrating the dependencies between cost features and other engineering information entities are to be defined and some elementary and non-exhaustive relations are shown in figure 3.1. Cost features, supported with a template library, is a conceptual class of objects that has to define the road map for doing the cost estimation within a specific scope. It has to define cost constituents, searching mechanisms and target elements in the scope and a sound cost structure. To interact with other types of features which are used for constraining the cost elements and rules, a unified feature scheme can be used. Such a unified feature system model covers a multi-face feature-oriented platform which defines the basic referencing mechanisms and information inference scope which can include concept features, machining features, user-defined face-cluster features, design features or any other kind of features.

Also, it has to define the level of details that is necessary for features and the structures that it wants to keep the feature dependency information. According to the scope and level of features, then the cost feature can be defined by identifying appropriate attributes and a set of constraints. Next, the tangible and intangible data sets are to be associated by identifying their various referencing sources, for instance, from CAD/CAM data, product configurations, ERP data tables as well as references to licenses, product models, historical data and suppliers' data. From CAD/CAM model and product configuration, design features can be determined. A design feature includes all the information about the material requirement, functional assembly features, functional part features and part geometry constriction features.

On the other hand manufacturing costs can be determined from machining features and machining processes. Machining features includes all the information about the machines that are to be used to produce the specific product, such as a waterjet machine, surface grinding machine, a welding machine, CNC turning or milling centers. Their specific data sets can be easily searched and used such as welding speed, equipment hourly rate and CNC programming capacity.

The most important step following is the mapping of cost feature attributes with design and manufacture features and testing their cost effect by using a unified weighting scheme. Then the variations of the effective cost features can be fully defined and the total product cost and the price quote will be determined. At the end, all the data need to evaluate the sensitivities towards changes of context and conditions as well as the reasonability of cost estimation model for different cases.

3.4 Feature-based semantic cost modeling

A real industrial case of oil and gas equipment was selected to describe feature-based semantic cost modeling. A *Tong* is commonly-used equipment for holds and provides the torque to the drilling pipes on an oil rig. Figure 3.3 shows a *Tong* body assembly in a 3D CAD model. According to the industry's historical data, the *safety door* is a critical part of tong body. Because the safety door is an important part of the product that the end users will touch to open and close during drilling operation and to maintain the tong body. The safety door is not a complex part but is welded together from six plates. Figure 3.4 shows the elementary plates of the safety door weldment, and table 3.1 shows the cost of each part that was determine in manufacture level.

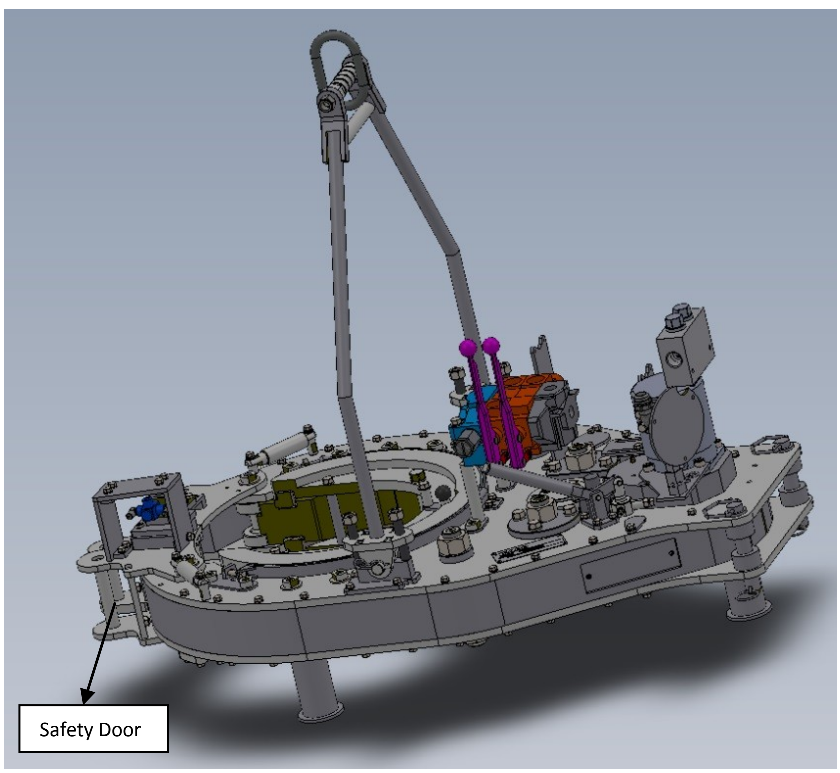
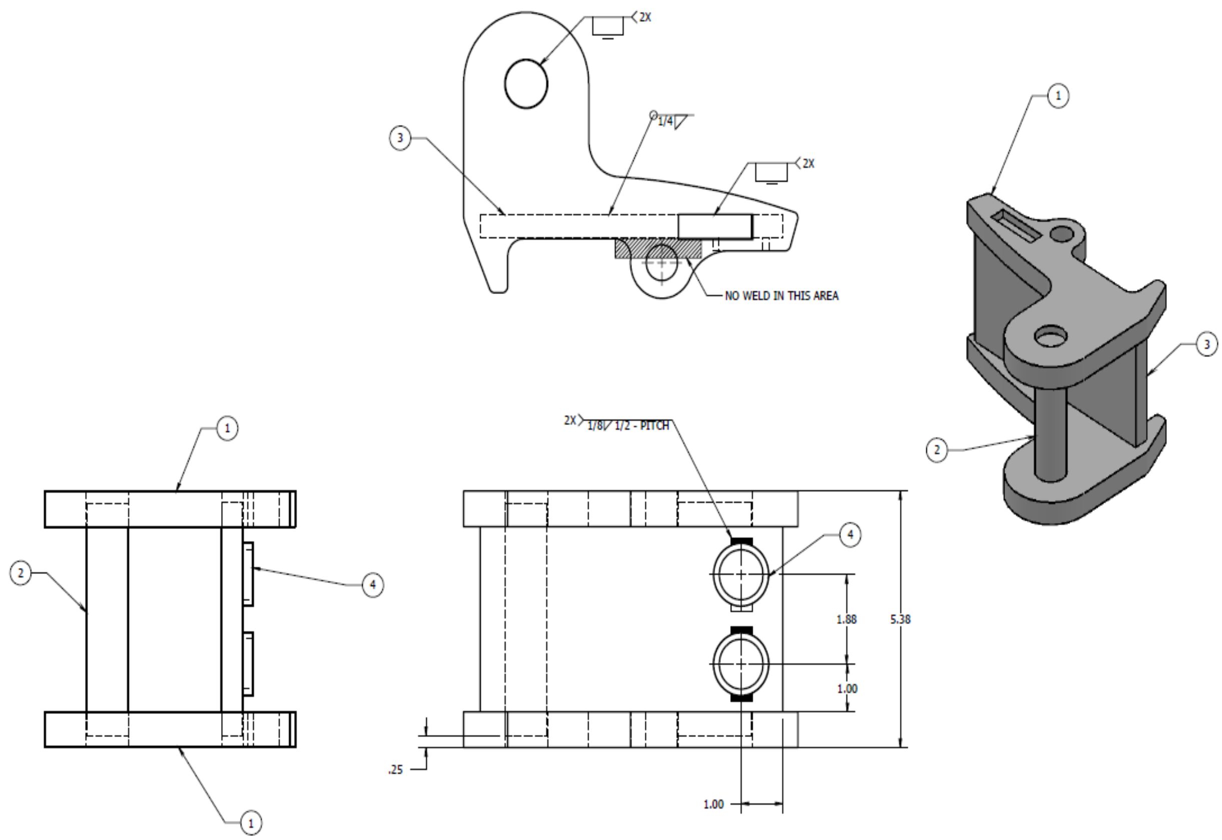


Figure 3.3 Example Tong Assembly Body 3D CAD model [86]



Legend:

① Latch (x2) ② Hinge (x1) ③ Spacer (x1) ④ Holder (x2)

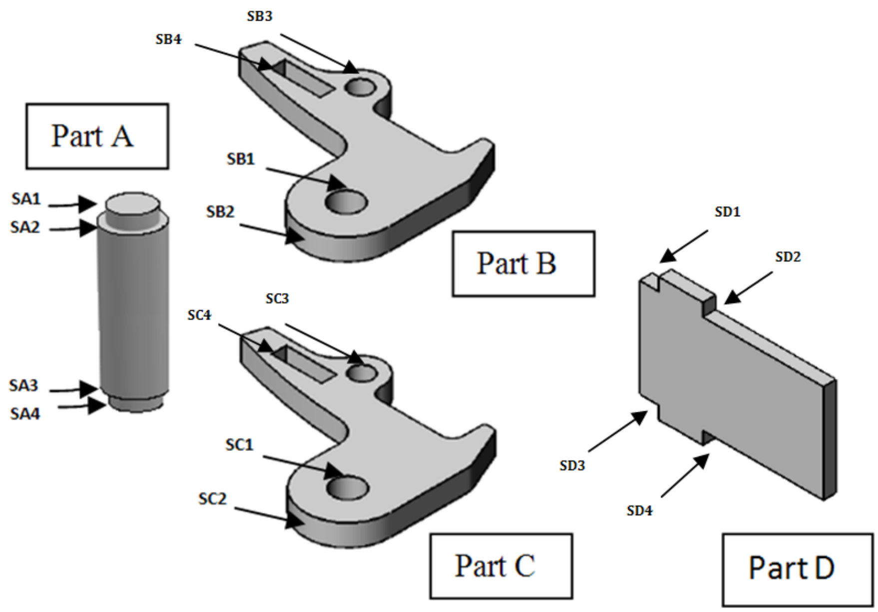
Figure 3.4 Safety door model [86]

Table 3.1 ERP information about safety door part [86]

Item	QTy	Title	Material Cost ($)	Labor Cost ($)	Burden Cost ($)	Time Spend (min)
1	2	Latch	24.17	15.7	5.6	30
2	1	Hinge	8.7	12	1.14	20
3	1	Spacer	27	8.4	2.17	12
4	2	Holder	11	3.84	0.7	3

In the current costing system of the company investigated, the ERP information shows the material, labor and burden costs for each item. However, there is no information about feature cost from the ERP database. In this case, if the company wants to change any of the design features, there is no systematic method to estimate the cost with new feature.

For describing our new feature-based semantic cost modeling method, at first, the cost feature has to be recognized. Figure 3.5 shows the parts and the face elements of the cost feature for the safety door. For example SD1and SD2 in Part D can be produced by cutting away the extra notches from the big blank part or welding the small blocks on a smaller blank prepared; both options can be evaluated from the input data extracted from the machining feature properties and using the relevant cost feature functions defined in relationship to the critical face elements.



Legend:

SX#: Surface # of part X

Figure 3.5 Cost feature elements for the safety door module [86]

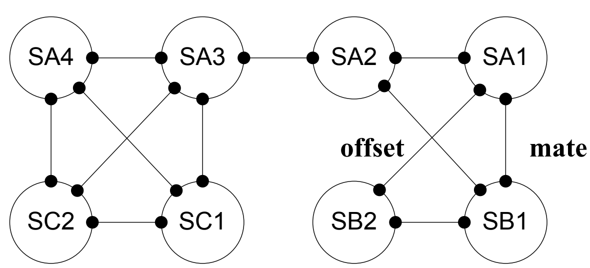


Figure 3.6 Interfacing feature faces between parts A, B and C

Table 3.2 Cost feature information

Part	Face	W_i	Thickness (in)	Min. feed rate (in/hr)	Quality	Max. feed rate (in/hr)	Quality
A	SA1	3	1.2~1.5	50	Good	75	Fair
A	SA2	4	1.2~1.5	30	Good	50	Fair
A	SA3	4	1.2~1.5	30	Good	50	Fair
A	SA4	3	1.2~1.5	50	Good	75	Fair
B	SB1	3	0.75~1	50	Good	75	Fair
B	SB2	3	0.75~1	30	Good	50	Fair
B	SB3	1	0.75~1	30	Good	50	Fair
B	SB4	2	0.75~1	10	Good	25	Fair
C	SC1	3	0.75~1	50	Good	75	Fair
C	SC2	3	0.75~1	30	Good	50	Fair
C	SC3	1	0.75~1	30	Good	50	Fair
C	SC4	2	0.75~1	10	Good	25	Fair
D	SD1	2	0.5~0.7	30	Good	50	Fair
D	SD2	2	0.5~0.7	30	Good	50	Fair
D	SD3	2	0.5~0.7	30	Good	50	Fair
D	SD4	2	0.5~0.7	30	Good	50	Fair

Figure 3.6 shows that SA3 and SA2 have a critical relation with each other as a result the weight of this relation is higher than the other features. SA2, SA3 is a hinge of control

for safety door. According to the interfacing feature relation schema, we can define the weight as an attribute for the relevant cost feature [63].

For example the weight of SA3 (W_{SA3}) is 4, because it has 4 connections with the other feature. There for, any changes on SA3 has cost effect on 4 more features, however any changes on SB3 just has an effect on 1 feature.

3.5 Conclusion

In this chapter, a new feature-based semantic model has been proposed for cost estimation purpose. This model is built on top of three sub models: feature-based association mapping, data mining and semantic modeling. Feature-based mapping model is used to determine feature scope and cost level defined, including all the dependency relations with other domain features such as conceptual, design, machining and user defined face cluster features. Data mining was proposed to identify correlations among tangible data, intangible data and human inputs for cost estimation purpose. Finally semantic modeling used to map the cost feature attributes with other predefined features, such as design and machining features. The chapter 4 will focus on mechanism implementation and full details of the multi-face feature models with the proposed automatic cost estimation model.

Chapter 4: Semantic Interactions within a Unified Feature Model for Product Cost Estimation†

4.1 Design model

Refer to figure 4.1, in this research, the suggested design model is feature-based that includes four sub-models: material requirement features, product assembly features, functional part design features and part geometry manufacturing features. The design model is created in the first phase of a product life cycle and enhanced or modified throughout other downstream stages including manufacture stage. It can be appreciated that the design sub-models product assembly features and functional part design features have a significant role in cost estimation. The product design model is detailed gradually based on the model's analyses by the way of further defining the model functions and purposes of the product. Along the evolvement of design model, engineers should consider processes of manufacturing the product. Generally, each design cycle has three main stages: conceptual design, preliminary design and detail design [64]. Conceptual design is about definition of product requirements functions, performance and possible design solutions. During preliminary design a few solutions can be selected for more analyses. Finally, the best solution will be selected at detail design stage.

There are several aspects for a well-defined design model, such as design for manufacturing, assembly, lifecycle, recycling, cost and etc. [65]. Design model also needs to present and incorporate the impact of manufacturing methods; hence design for manufacturing (DFM) sub-module is necessary. Ideally it is a systematic approach that considers the available facilities, tools and capacity in the stage of design product with

† A version of this chapter has been published in International Journal of Mechanical Engineering and Mechatronics. Sajadfar, Narges, Luis Campos Triana, and Yongsheng Ma. "Interdisciplinary Semantic Interactions within a Unified Feature Model for Product Cost Estimation."

features so that the product can be simply and economically manufactured. DFM is the synthesis practice for designing component functions and strengths with additional considering elements according to the manufacturing requirements. DFM module in a computer integrated system describes operations such as welding, cutting and milling for each component. For example, for a welding operation, it needs to describe the position, height, length and weight of welding and all the drive path profile information. Note that, the design model, based on the machining model, can enhance the product accuracy cost effectiveness, and manufacturing flexibility. However, even in such a comprehensive design model, it does not have enough input to include real-time machining information, such as actual machine allocated cutting conditions, e.g. the feed rate. The designing for cost is a common demand in the purpose of reducing product cost, but so far many companies have difficulty to organize and extract the useful data for this advanced application aspect.

The contents of design features change constantly during design process and so does the development of sub-models. Therefore, associative design and manufacturing features need to be used [58] to keep the consistency. The designer can use advanced feature-based technique with a persistent design feature representation during the whole design process. However, it is necessary to distinguish that design features are not interchangeable with manufacturing features [63]. Figure 4.2 illustrates the model and two views of a key component in an oil drilling equipment manufacturer, referred to as *door weldment* as a real case study. Table 4.1 shows the member work piece information of the *door weldment*. In concept design model, each member work piece in the drawing can be defined as a weldment design feature. For example during conceptual design *door spacer* is defined as a side impact protection feature. During the detail work piece design, then the exact location, size and material of them are defined; and then some material cost related data can be worked out.

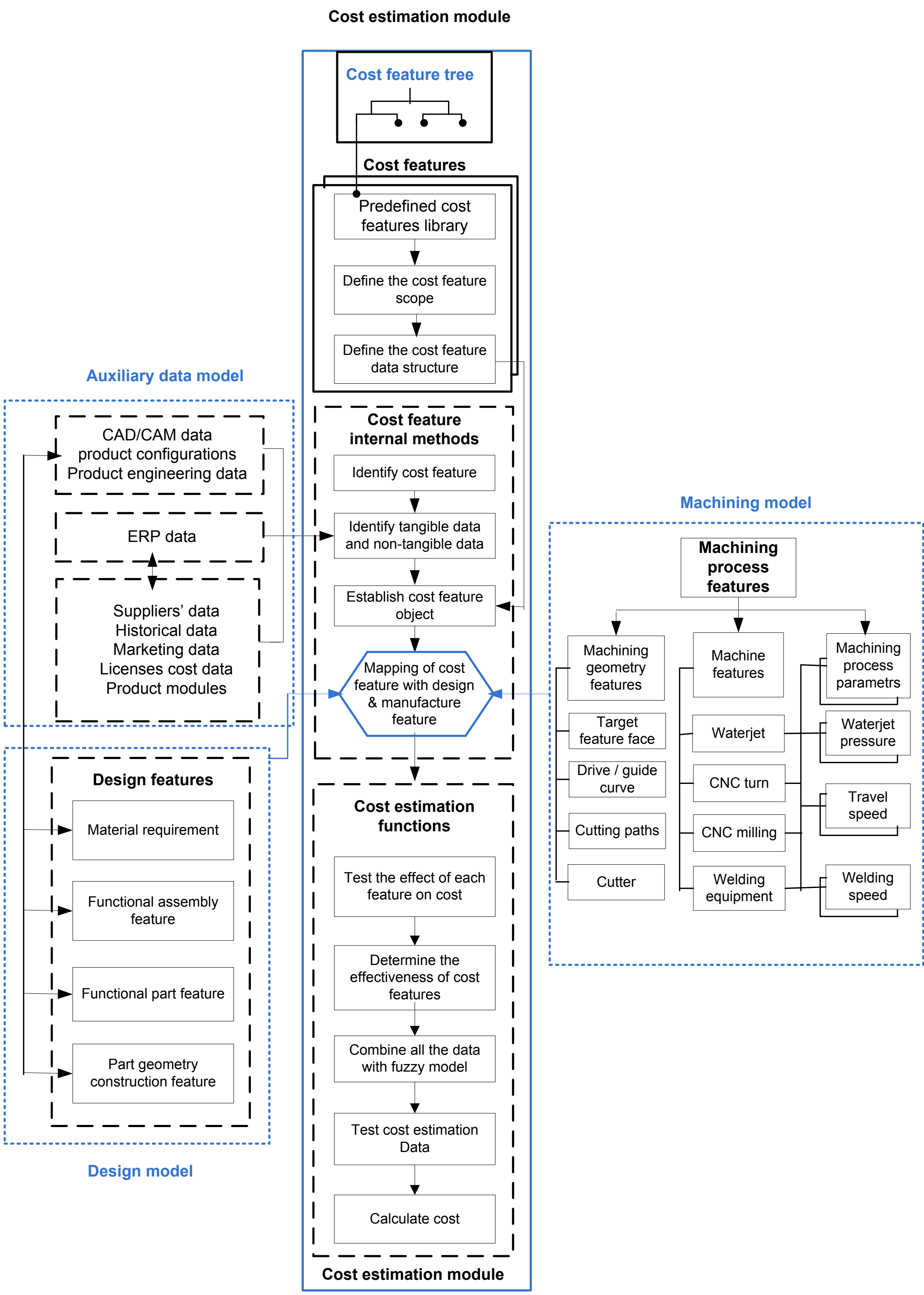
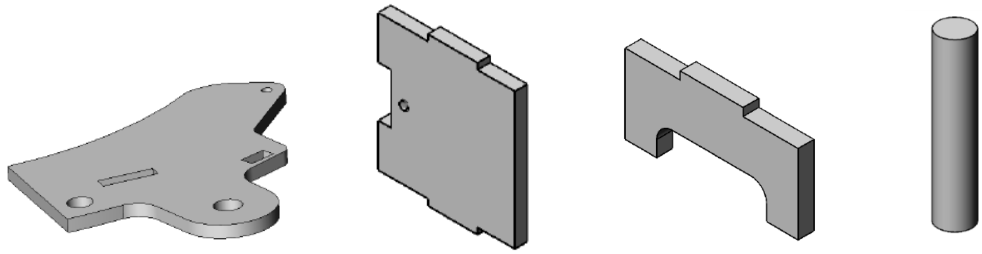
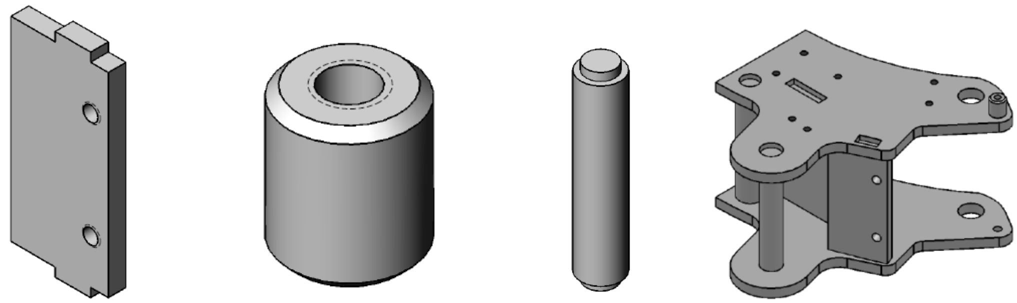


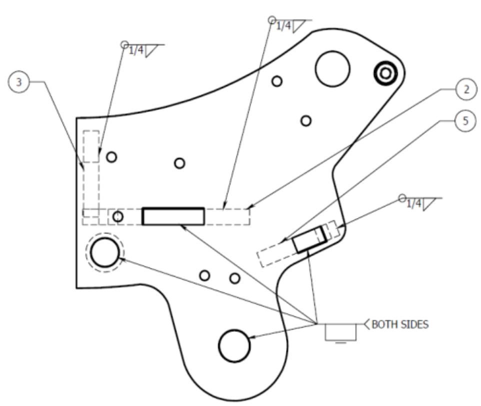
Figure 4.1 Cost feature semantics with non-exhaustive relations



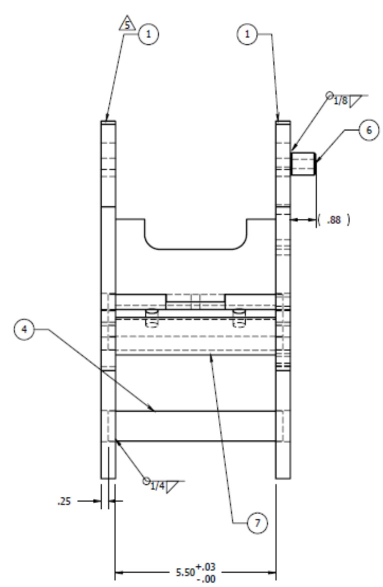
1) Plate door right hand 2) Plate door spacer 3) Plate door spacer 4) Post door



5) Plate door spacer 6) Door cylinder 7) Post door - Door weldment assembly



- Top view of weldment drawing



- Side view of weldment drawing

Figure 4.2 Assembly drawing of door weldment and its individual work piece [86]

Table 4.1 Detail design information for work pieces before welding

Item	Qty	Work piece name	Purpose	Weight (LB)	Material code
Top & Bottom plate	2	Plate	Plate door right hand	9.61	Medium carbon steel
Spacer	1	Plate	Plate door spacer among Top & Bottom plate	4.27	Medium carbon steel
Spacer	1	Plate	Plate door spacer	1.54	Medium carbon steel
Post door	1	Post	Post door	1.34	Medium carbon steel
Spacer	1	Plate	Plate door spacer	2.18	Medium carbon steel
cylinder	1	Lug	Door cylinder to make a hole	0.09	Medium carbon steel
Post door	1	Post	Post door	2.01	Medium carbon steel

4.2 Manufacture model

The suggested manufacture sub-model in this research is also feature-based. Each manufacturing process is embedded into feature which can be initiated from predefined templates and fully generated with extracted feature entity information. For example, machining feature associates the blank geometry, the related target part geometry, the cutting geometry; the machines used, and detailed process conditions. This manufacture sub-model has to represent the involved process data and entities via features; also, it has to manage the relationships among the feature entities and the process data. Refer to [63] for more details from both theoretic and application aspect.

Figure 4.1 provides the breakdown of the manufacture model, which has two main parts: manufacture features and manufacture process. For example, a machining feature's geometrical entities are generally defined as a set of faces and while it's related tool path driving curves are defined by referring to the part geometry as a set of derivatives; the part and the derivatives are represented consistently with the Boundary Representations (B-Rep) geometric model. Manufacture feature geometries are defined according to associative features encapsulating the characteristic parameters of specific machining processes, such as typical holes, slots, pockets, that can be identified from the model of the final product [63].

There are several approaches for manufacture feature reorganization such as graph-based, volumetric decomposition and hint-based [63]. Many researchers also have attempted to define a standard set of machining feature, such as the Standard for the Exchange of Product Model Data (STEP) that has been an international standard for representing engineering data structure and exchanging product data model [58]. Theoretically, with the generic feature concept and further research work, user defined machining features can be incorporated as well. However, cost information data structure associated to design and manufacture features are not well defined so far. Table 4.2 shows the cost elements related to some manufacture processes for a *door weldment*. Five different operations are needed and the machining features costs should be linked with associated defined elements. Machining feature cost and its elements will be explain in section 5.4 , the definition of welding feature as a case study of machining feature will be describe with more details in section 6.1 and 6.2.

In addition, the machine configuration information such as equipment type, power, dimensions, allowed interpolation, access and traversing directions, rotating, and positioning attributes of each machine are modelled in the form of machine features. Machine cost has to be considered as well.

Table 4.2 Manufacturing Information for the door weldment.

Operation sequence	Operation code	Operation description	Work piece name	Run	Run type (hours or minutes/ piece)	Setup cost ($/per hr)	Running cost of the machine ($/per hr)
1	Water jet	Cutting	Top & Bottom plate	1.3	Hrs/PC	$30	$60
2	Man-Saw	Cutting	Spacer	6	Min/PC	$30	$60
3	Drill & Tap	Drilling	Spacer	4	Min/PC	$20	$50
4	Man-Lathe	Turn/Face	Spacer	9	Min/PC	$30	$60
5	Weld	Welding	Door Weldment	3.2	Hrs/PC	$10	$25
6	Mill	Milling	cylinder	0	Hrs/PC	$30	$60

4.3 Auxiliary data model

Auxiliary data model is a collection of "design for x" data model under the concurrent engineering concept, which supports product development and manufacturing aspects with certain specific engineering and optimization. This model can be recognized as a data association system that enables the sharing and updating mechanisms of system integration across the product life cycle stages. For example, lots of tangible data need to be associated for product cost estimation such as CAD/CAM data, product configurations, profiling information, ERP data, available machines, supplier's process and cost data, and product quality models. All of these data sources and uses need to be identified at first step of auxiliary data model. Product feature library can define the commonly used classes or objects information as templates to gather cumulative enterprise knowledge and all the associated information accurately.

To introduce the proposed cost engineering feature model, a *cost feature* definition is hereby given: *Cost feature* is a class type under object-oriented software engineering methodology, representing an abstracted constituents of costing items in the form of commonly recognized patterns, which contains the characteristic properties and behaviour functions. The properties include characteristic attributes, geometrical entities (including sub-features), referenced entities (including other features), etc. The supporting methods of the class include creating, managing and deleting predefined relations among the member entities, evaluation and validation methods of the attributes and entities of the class; and the expected behaviour functions according to the generic needs of cost engineering applications. Then individual *cost features* as objects can be instantiated from the *cost feature* class to materialize individual applicable cost items.

Given the variety of cost items to be considered in engineering, cost features can have different variations and they can be accommodated by applying object polymorphism techniques. So are the cost feature member entities and attributes, specific sub-types can be defined for different manufacture process categories

The main idea in this research is that cost feature member attributes and entities can be identified and modelled with the help of data mining (DM) technique. To start the description, the scope of *cost feature* data mining approach must be defined. It should be started with the investigation of the scope of data mining applicable in this research field.

As we have discussed before, there are many sources of tangible data for cost engineering. For example to set the price for an existing common part, we can usually collect and refer to all the market prices. However for a brand new product, there is not any market price and we have to look around for those similar products as well as the benchmarking prices about the new product, and create a logical pricing formula the reference to the elements of the auxiliary data model. However, product pricing is very much dictated by the manufacturing costs as well, and they need to be worked out first too. Hence, eventually cost features are useful for cost engineering throughout a product's life cycle. To do that, different levels of information from different angles are needed;

hence, it is necessary to define product cost feature from deterministic relations among all the related elements of any product in sales from raw material costs to delivery costs.

However, the proposed DM approach is based on another kind of data, which is intangible data, which means the data has no clear (or too complicated) relations with the targeted cost items. This kind of data has indirect effect on the cost estimation, such as the computer software licences cost that used in design and manufacturing. It is reasonable to assume that, after gathering all kinds of data with the help of modern enterprise software tools, like ERP systems, CAD/CAM tools, or process planning tools, the related cost feature member attributes and entities could be discovered by data mining algorithms. Relating cost features with the relevant data set selected among the huge amount of data is a challenging task that needs expertise.

4.4 Data obtained by employed sub-model

As shown in figure 4.1, the auxiliary data model contains ERP data, historical order data with cost components, and product model configuration data. Obviously, useful data can be extracted from such information repositories for cost estimation purpose. For example design features can be analysed as they are commonly cycled in each product order. Detailed design features can be searched and their pointers collected in a set of attribute data structure because they have been fully implemented in the CAD/CAM data model when the products were developed according to design feature templates. Further, by searching the manufacture process model, the predefined process features can be listed and tracked. However, to extract cost data from the ERP and order data models to establish the cost feature objects, the cost feature templates, more research often has to be applied. To do so, the cost feature library needs to be defined with the cost feature data structure and the data extraction methods are required. It can be appreciated that typically the three data models contain only past order data with those related design and manufacturing features while the explicit cost relations are not yet identified.

To solve cost feature identification problem, the historical order data with cost components from ERP data model has to be analysed and associated dynamically to the design and manufacturing features in order to justify the estimations and also directly support the functional estimation module. Therefore, the next step is to develop the feature mapping relations based on the results of the above data extraction procedure and identify those input relations that will be used as cost feature mapping technology.

Up to now, several methods were introduced for feature mapping, such as mathematical model [66]. To keep the implementation simple, the mapping matrix method suggested by Zhang et al. [67] is used for implementation in this work. The outcome of mapping matrix can illustrate the cost estimation function. Mapping matrix is sufficient to manage and transfer different cost features, however it cannot cover all the relation types or details. By using the output function generated from the mapping matrix, the cost features are identified for each product cycle; and hence the accuracy of cost estimation will be increased.

4.5 Data mining for cost feature pattern construction

Clustering and classifying the data can be useful for data management. The first step is clustering the available data according to application scopes. This step can be done in two ways: clustering and classifying according to the assembly principles of component analysis, or using a predefined clustering algorithm. Then, within each cluster, DM process can be used for further cost engineering. By applying DM in a cluster with the appropriate size, the accuracy of data mining results will be increased, because there is a set of common relations among the components of each product family which has similar data patterns. Within a cluster, components will be classified by the same decision tree and rules.

The second step is to classify cost related data among components in a cluster. During the DM classification process, two kinds of data will be encountered: quantitative data and qualitative data. Quantitative data can be classified by a decision tree via rule setting. Linear regression, K-nearest neighbour and baseline predictor are some common algorithms that can be used for analysing qualitative data.

In the next step, by analysing the combination of classified quantitative and qualitative data, the cost model patterns will be extracted. Model patterns can be used to filter the selected cost related data and as a result the cost feature constituents can be established with validated DM processes. At the end, these constituents, in the form of attributes or pattern data structures are clearly identified and created into reusable class elements. These elements are then further ranked according to the nature of the product that the cost estimation method will be applied. Also, there are several data mining algorithms reported for data ranking. Such as, best first (find the best data first), linear forward selection and ranker (find the best set of data), which can be useful for data ranking. Then, one of the most effective ranking techniques can be selected based on the weight of data or accuracy of data [68].

Figure 4.3 shows the steps of auxiliary data modelling with more details. Table 4.3 shows two resulted patterns generated based on cost related data that can be used for the estimation of cost for producing *door weldment*. Note that the simple average cost per part accounting all the cost elements should be $831 for alternative 1 while $518 for alternative 2. However, the resulted average cost per part after data mining is not exactly the same. They are $787 and $483 respectively. The differences reflect the influence of intangible data on the final cost due to the inconsistent data sources and discretion by the management. The interpretation and accuracy evaluation for the data mining method are to be further studied in the future. The system design details are to be introduced in the next section.

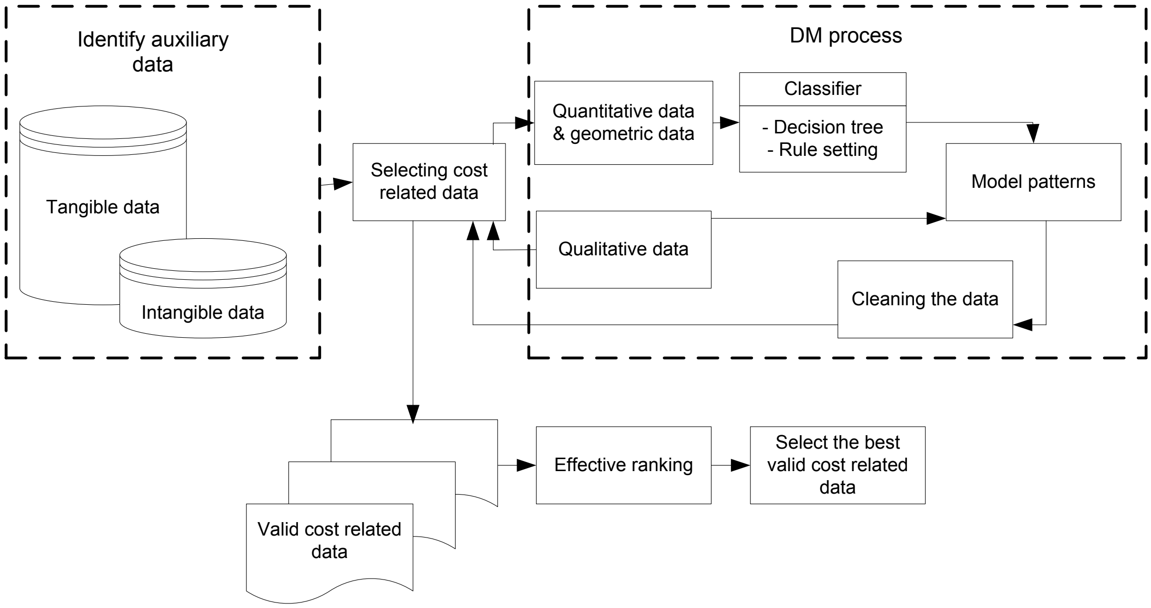


Figure 4.3 Auxiliary cost modeling with data mining process.

Table 4.3 Generated output by integrated model.

DM method	**Machine cost/yr ($)**	**License cost/yr ($)**	**Maintenance cost/yr ($)**	**Labor cost/yr ($)**	**Material cost per part ($)**	**Output part/yr**	**Average cost/part ($)**	**Quality**
Alternative 1	420,620	25,000	420	1,600,000	45	2600	787	Good
Alternative 2	320,000	25,000	360	1,200,460	35	3200	483	Fair

4.6 System design for cost estimation

As explained, having a unified cost estimation system is imperative for doing dynamic cost engineering throughout the stages of a product lifecycle. Figure 4.4 illustrates the computer system design for cost estimation conceptually with reference to the modules of an ERP system in use in the sponsoring company and those related product engineering models. Cost estimation system begins with the customer ordering process. The *customer requirements* are first collected, processed and recorded into the ERP customer relation management (CRM) module. Then such requirements are mapped into product *functions* which are defined by the product model series, and up into a customer specific *product configuration* [50]. Based on such a *product concept model*, *unified feature model* is established which includes *design features*, *manufacturing features* and the corresponding *cost features*. Then, with the relevant information provided from the product model, a *cost advisory system* module needs to manage the cycles of "if-then" design analyses about the costs. Via the cost advisory module UI that shows in figure 4.4, the interactions between customer requirement and cost engineering, modules which are supported by the organized cost features, can be realized dynamically.

From the synthesized cost results which are functionally supported by the cost advisory system, the customer is then informed on the estimated product cost much more accurately than using traditional experience-based rough estimation. After, cycles of evaluation on the product in contrast to the cost specifications of the product expected, the customer can make then the purchasing decision; and a customer order is generated.

Once the customer order is confirmed, from here the cost management process kicks in. The cost management system is activated and the corresponding UI as described later in this chapter shows the actual cost components of that customer order. In the next step, all the cost information from *cost advisory system* will be transferred to cost management module and it will also be used in the ERP system. Cost features are used in mapping cost related associations by referring design features and manufacturing features under the product unified feature model.

4.7 Integration among sub-models

Each of the above data models can be defined as a module in the system and mapped by a functional modular class; i.e. design class, machining class and auxiliary data class. Semantic data modelling between these three classes provides the integrated system within different views on data; such as structured and non-structured data [69]. To realize the expected functionality, several types of class relations among sub-models can be defined. A relation can be defined as a semantic link between two or more objects for the purpose of creating a logical system. Conceptually, a relation can be defined as an attributes, entity, independent element or function [70]. In the sense of object-oriented approach taken by this research work, the relations among classes can be defied as an aggregation, inheritance, using, association, and instantiation [71].

Figure 4.5 shows the UML diagram of integration with sub-models. Cost features are associated with all the three sub-models, or the three modular classes. First, there is an association relation between design feature class and machining feature class. Due to the interlocking nature of design and manufacturing considerations in real product lifecycles, there is some kind of mutual dependency link, *cooperate*, between these two classes. Auxiliary data model is an abstract modular class that has "*reference*" relationship from the design feature class and the manufacture feature class. It means *auxiliary data* class is dependent on *design feature* class and *machining feature* class; and any change on the design and manufacturing will be automatically reflected on auxiliary data class properties. Cost features refers data input from all the three sub-models; but the real relationship materialization can be constructed based on the input data and the data mining process results as discussed in section 4.6. Therefore it is clear that cost features are modelled to take care of the intricate cost influencing relations initiated from design and machining features as well as other auxiliary data.

It can be appreciated that usually, the features that used for design have different definitions from those related manufacture features. To clarify the interdependency relationship between design and manufacture features, more discussion is necessary. Figure 4.6 shows a part of *door weldment* and their geometrical elements of design features and the machine features on this part respectively. In design geometry, the part was defined by adding a small block on top of base block and cutting a curved notch from bottom of base block. On the other hand, the part machining process was defined by water jet cutting represented with machining feature profiles B, C and D.

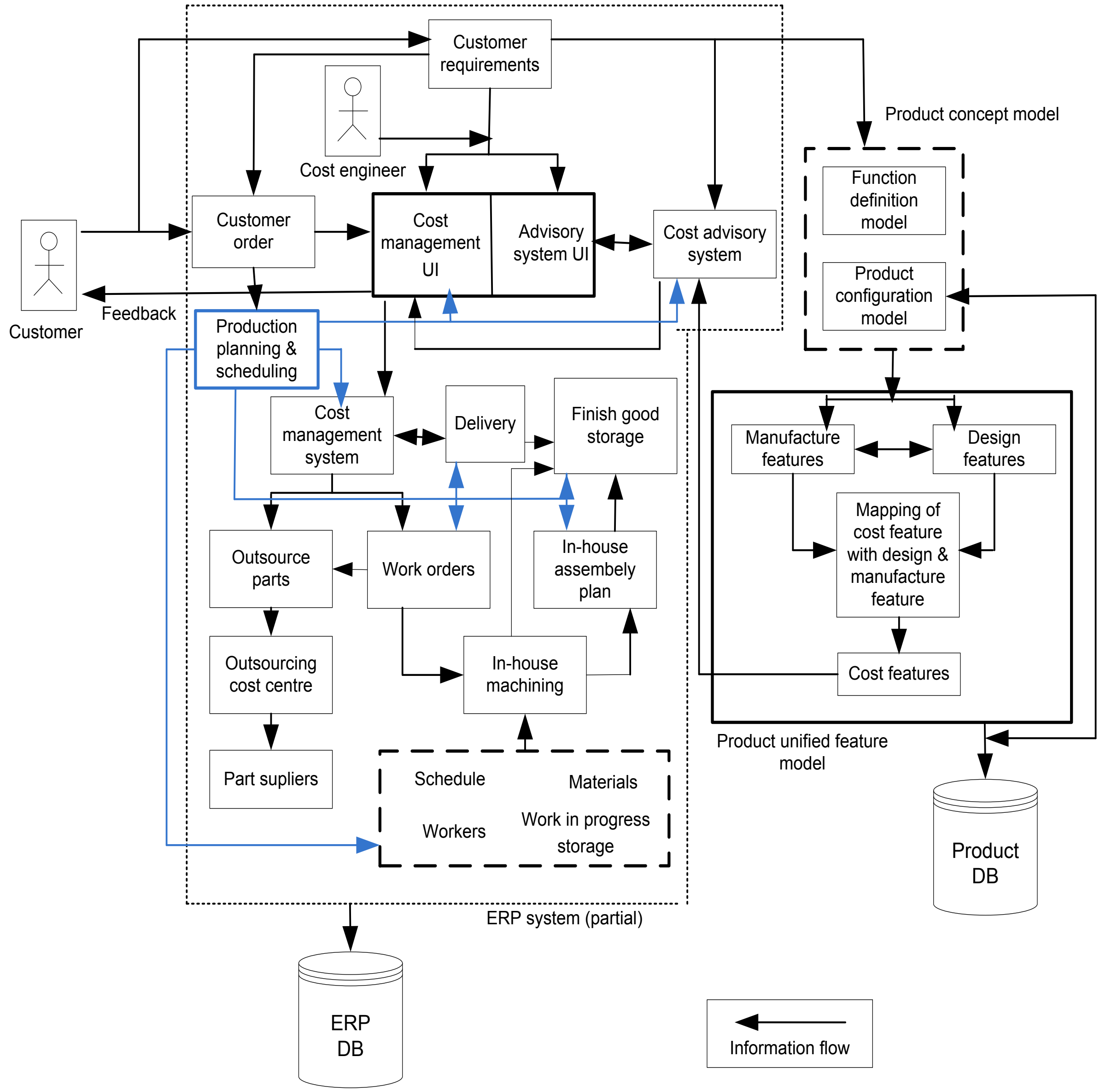


Figure 4.4 Computer system diagram for cost estimation

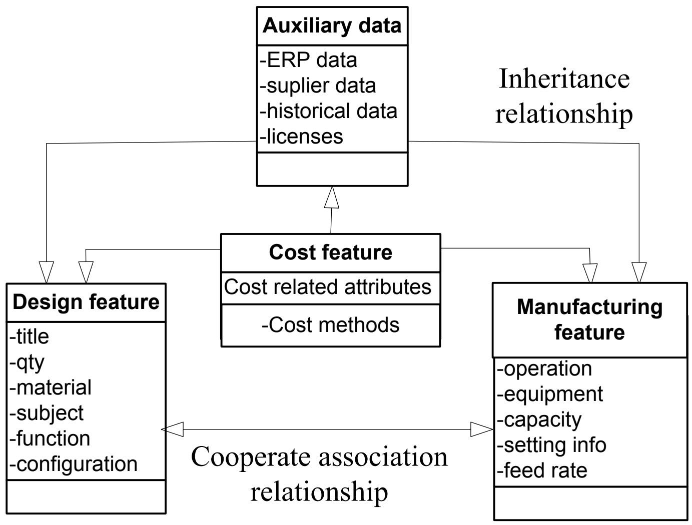


Figure 4.5 UML diagram of integration between sub-model features.

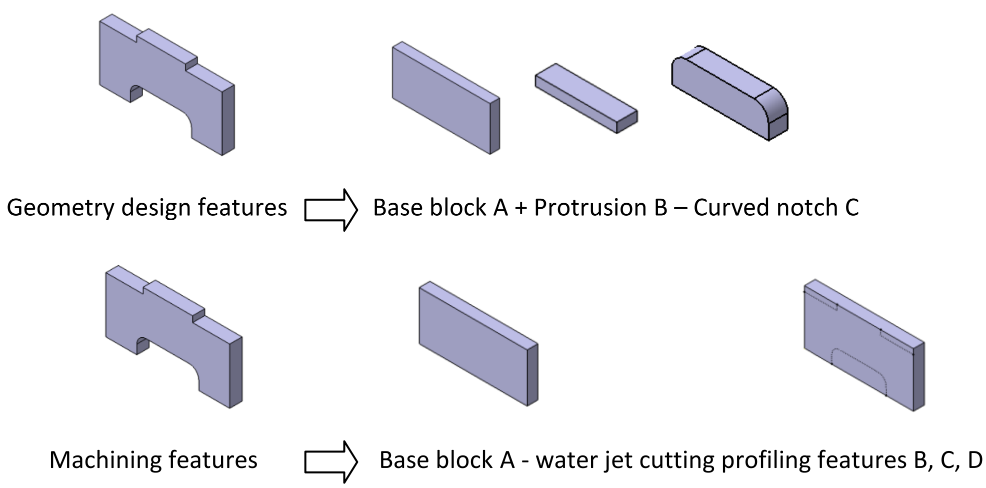


Figure 4.6 Difference between machining features and geometry design features

4.8 Conclusion

In this chapter, the author defines a *cost feature* class, and then presented an overview of the relations between cost features and three engineering sub-models, i.e. machining model, design model and other auxiliary data model. For having the dynamic, unified and accurate cost estimation modelling, the relations among cost features and all engineering sub-models need to be managed consistently. The descriptive semantics of cost feature is flexible enough to represent all types of associations that are related to cost estimation, having deterministic and non-deterministic. For the latter type of relations, data mining is proposed as a useful method to extracting data patterns for these three sub-models. It was also discussed that clustering before data mining is helpful to assure the quality of the approach. Classifying cost feature elements according to data mining algorithms was proposed for intangible data analysis in the purpose of more accurate cost estimation. Currently, the data application and error evaluation have not been carried out yet. They are two important aspects to proof the proposed algorithm; and they will be addressed in the following chapters.

Chapter 5: A Hybrid Cost Estimation Method Based on Feature-oriented Data Mining ‡

5.1 Current practice of cost calculation in ERP systems

Due to the increasing manufacturing product complexity, variations, as well as dynamic supply chain management, manufacturing companies need to coordinate its order acceptance process dynamically hence they must make decision based on timely information resources among their business partners. The customer inquiry management has to be done accurately and quickly to provide effective feedback.

Currently, different computer-aided systems are used for collaborative manufacturing that aims to increase efficiency and flexibility while also keeping sensitive collaboration loops with the different business partners. Enterprise resource planning (ERP) is one of the most important management order-fulfillment tools capable of creating the links between product documents, schedules, and other forms of communication. ERP software is used to share information among different partners for daily manufacturing activities, through which many of the process functions are intended to support collaborative manufacturing, such as customer ordering, process plans, financial decision making, accounting, and supply chain coordination, all of which are unified and managed from a single system [72].

However, ERP packages currently do not have the sufficient functions necessary for market-oriented cost estimation with dynamic marketing and product configuration, even though such functions would play a major role in business success [73]. Cost estimation

‡ The major part of his chapter has been submitted to Advanced Engineering Informatics, by Sajadfar, Narges, and Yongsheng Ma.

(CE) is one of the key success factors that create integration between in-house business and market-oriented functions.

Manufacturing costs in many companies are not accounted for systematically and accurately due to the complexity and the constant changes of the processes involved in production and the lack of data collection schemes. The process model in an ERP package contains certain details of manufacturing processes, such as the types and setups of machining features, cutting conditions, expected productivity, past costs, and status of job completion. As a result, in this work the process information is integrated within ERP for CE purposes. However, one limitation of most ERP systems is that the actual variations of processes, such as the different welding jobs, are not well defined and documented. Such technical challenges are particularly true in companies with small batches and high variations of product models, which cause overwhelming difficulties in accurate CE. Developing a generic and comprehensive data model for manufacturing processes in ERP has been a major issue that is being addressed to achieve the expected production performance [73]. Most ERP systems can estimate process costs based on the required time for the machining process (i.e. labor cost), materials, and power consumption. However, usually, the complete information for a specific manufacturing process is difficult, and making use of the historical data from similar operations requires an expert judgment to determine the factors that correlate to the respective costs.

Traditionally, the cost of a product is calculated based on the time spent on product development, the manufacturing processes required, equipment used, and the energy and materials consumed. In turn, these are influenced by the parts' shapes and characteristics. On the other hand, the modern cost paradigm focuses on customer inputs, product similarities among configurations, and flexible markups [74]. Beside cost calculations, there is a considerable amount of research available on CE for manufactured goods, specifically about the manufacturing processes, such as turning and milling CE. However, many process costs have not been clearly defined yet, such as the cost of welding operations in variation workshops. Analytical and parametric methods are commonly used for manufacturing CE. Such methods are difficult to adapt to different jobs, and estimation can only be given by experienced personnel [73].

5.2 ERP solutions

ERP vendors are developing stronger financial functions for cost management. A typical ERP system records and tracks product and process related data such as raw material, set-up time, run time, number of machines and labor time for cost calculation. The disadvantage of this method is that many detailed features and parameters are not considered in cost calculations; therefore, the accuracy of the cost results is questionable. In addition, many cost items can only be calculated when all of the parameters are in place, this constraint requires more time and data or even makes timely cost estimation too expensive to be implemented

To solve such problems, ERP vendors are developing accurate cost forecasting methods. For example EPICOR™ [75] is working on the development of cost forecasting and planning in addition to making calculations of manufacturing costs. In the EPICOR ™ system, the budget process is defined based on the manufacturing process plan in multiple cost-calculation scenarios. Then, the financial planner uses different methods, such as examining history and trends and making a depreciation analysis, to do the forecasting [75].

Currently, in an ERP system like Epicor, the parametric CE is still the basic method for quotation generation and final cost determination for a company's manufacturing products and processes. Usually, the ERP system conducts cost calculation based on four categories: labor, material, burden, and service, which can provide the general information necessary to determine cost. To have a more accurate CE and calculation, more attributes needs to be extracted. As an example, figures 5.1 and 5.2 show the cost attributes of a specific part (Part ID: 1095-139 from studied company) in EPICOR and [76].

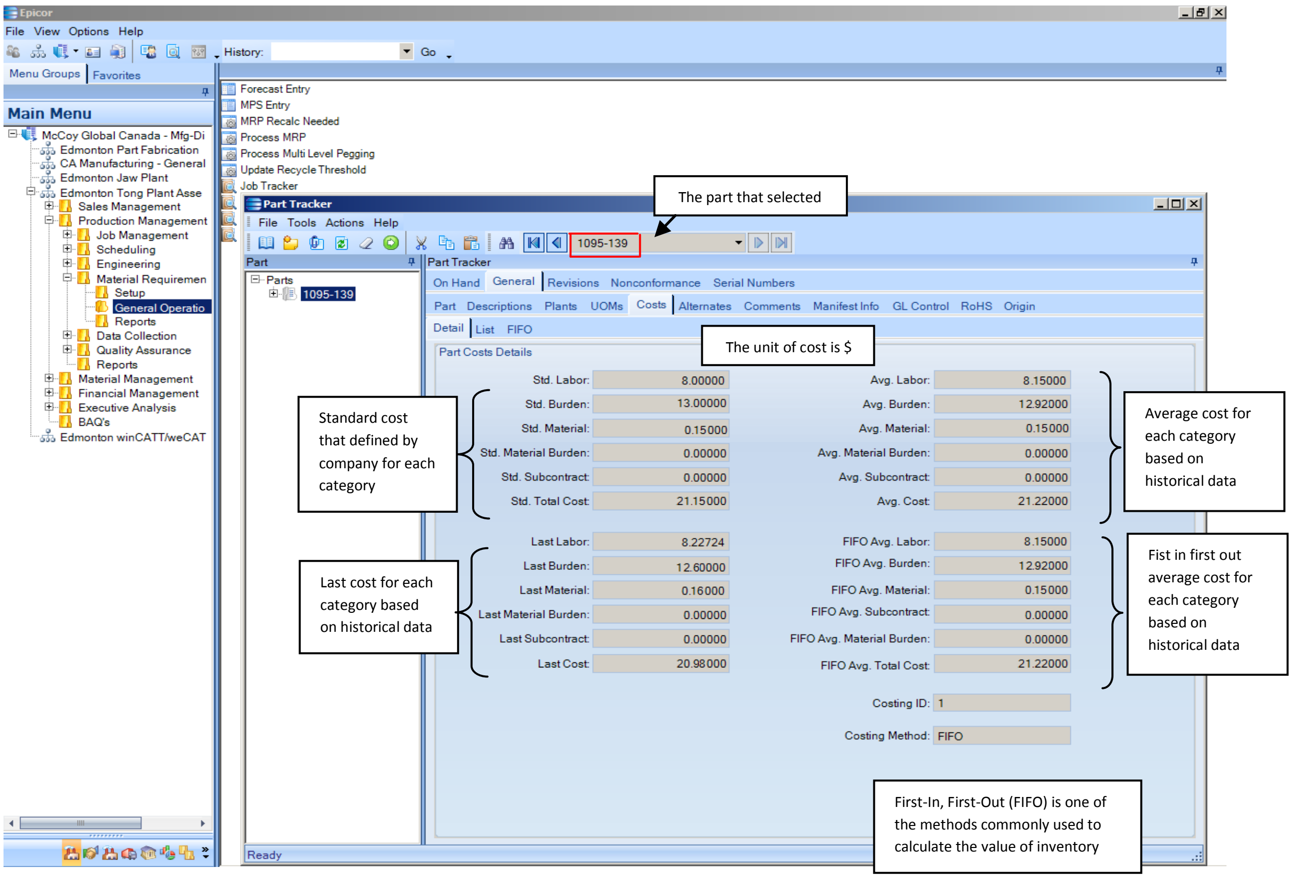


Figure 5.1 Typically-defined cost attributes in EPICOR™

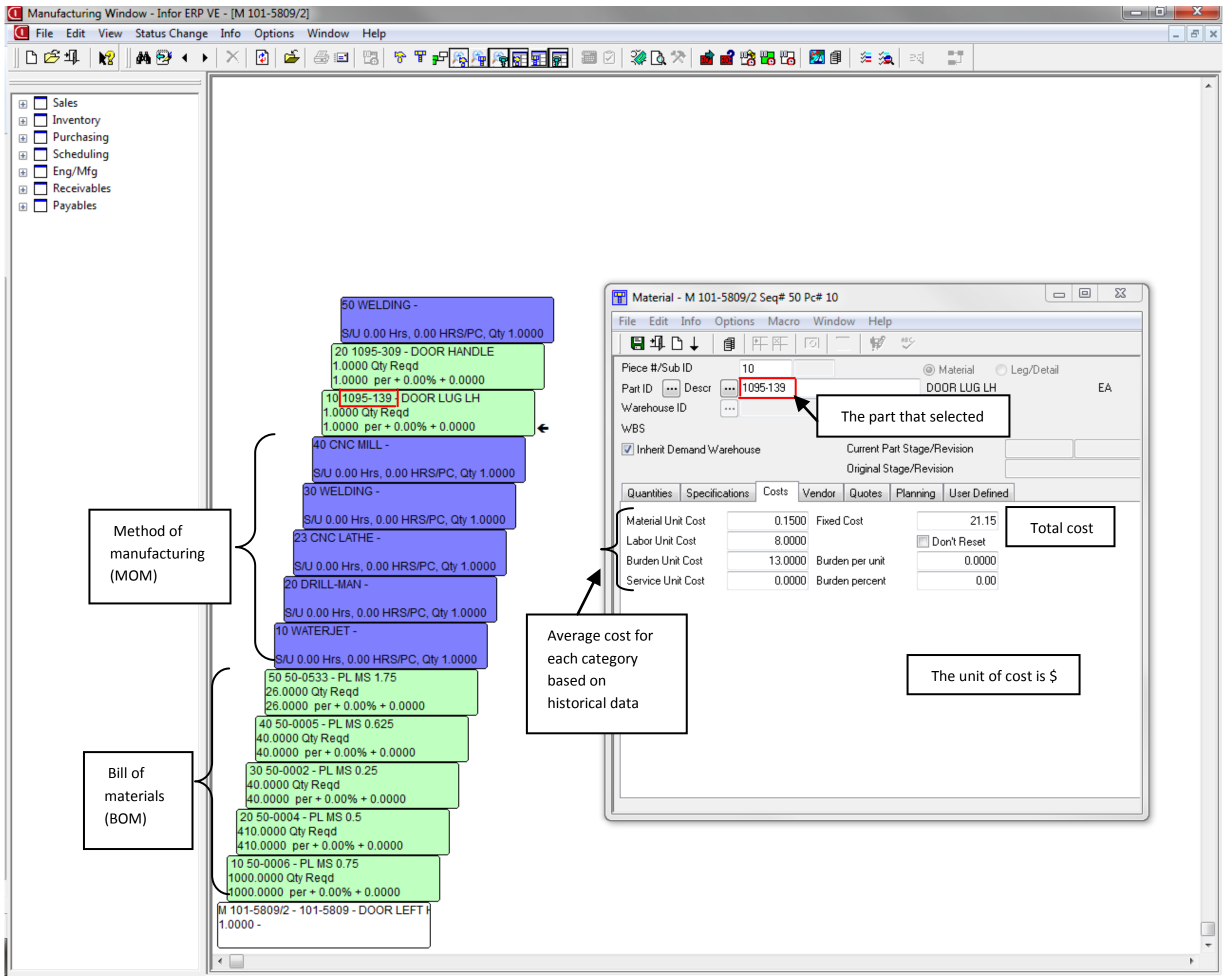


Figure 5.2 An example set of cost attributes defined in Visual Info™

5.3 The principle of hybrid cost estimation method

A hybrid CE method includes two main modules: empirical analysis and DM analysis. Historical data and cost feature libraries have been used as input for these analyses. The initial input of the CE system is the historical data, which is usually available in many companies but in different table forms. To unify the data structures and create a characteristic representation of cost components and calculation algorithms associated, a new concept of the feature-based CE model has been introduced by authors [58, 77]. The authors have defined *Cost feature* as a unique class in the unified feature modeling system to address the characteristics of cost-related data, restrictions, and dependency relations, which are closely related to the product configurations and the manufacturing processes involved. By using the product and manufacturing process information with the support of historical data information, cost features are extracted into consistent data structures, and templates are created to assist complete data gathering. They are organized into a pre-defined cost feature library, making it available for the feature-based CE. The cost feature class diagram is introduced in section 5.4.

With the sub-classes of the cost features, all components can be divided into different categories according to their principal functions, geometry information, similarities, and process plans. As a result, the feature library can classify common-feature families into the same class. It can also use a feature-mapping technology [66, 78] to create a classification for each manufactured product. The proposed method uses pre-defined feature objects to create links between each feature template (class) and its instances (details of costs for the data items).

In this proposed system, to support the hybrid approach, a CE engine has been constructed with the inputs from the historical data and the cost feature library, as figure 5.3 illustrates. An empirical cost analysis method has been developed based on the approach suggested by Masmoudi et al. [46, 47]. Regression analysis with expert judgment is carried out to develop a feature-based cost calculation formula. As figure 5.3 shows, an empirical analysis is one of the two complimentary CE methods to estimate product costs and their accuracy.

In contrast, in another branch of the hybrid system, the DM software algorithms are applied to the sample historical data to analyze and extract the pattern for each data set. After evaluating the cost patterns, the DM module will estimate the cost components and their accuracies for new product inquiries.

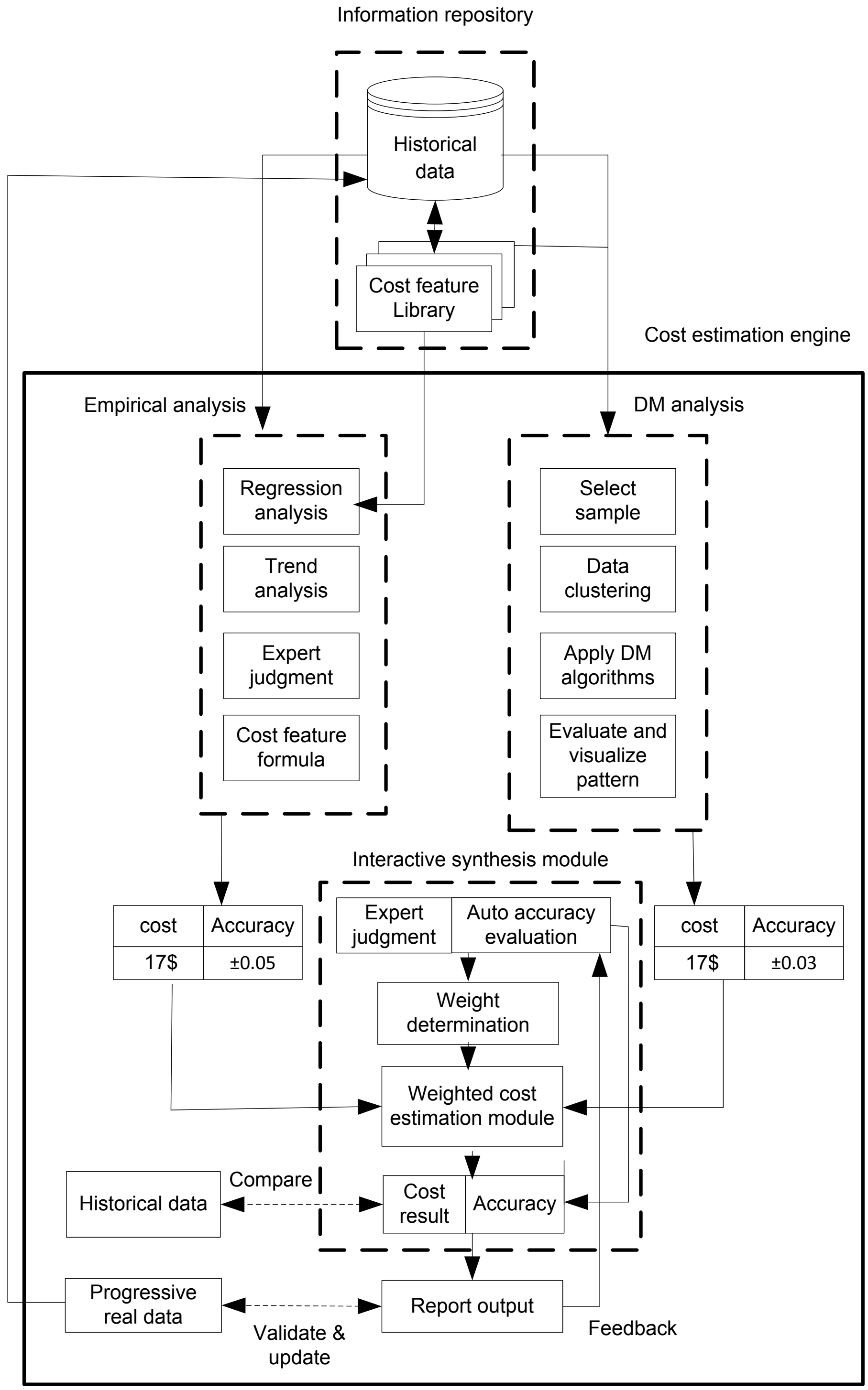


Figure 5.3 A hybrid cost estimation method proposed

These two CE methods are validated with the historical data, and they are combined with a weighted sum mechanism, as figure 5.3 shows. The weight given to each method is based on expert judgment from the validation results or is suggested by a system's accuracy evaluation algorithm. The authors believe that the proposed hybrid method can achieve accurate CE with adaptive adjustment capability in real implementation.

5.4 Definition of the cost feature

As described in previous chapters, a cost feature is a set of an object's characteristic attributes and their related functions and servicing methods, which can be associated and constrained to represent the pattern of semantics for the purpose of cost engineering. The cost feature has to consider the variety of cost-related information: quality; operation sequence and time; design; equipment; materials; functions of application; and geometry and non-geometry details, such as tolerance, surface finish, and working allowance. Also, each of these factors has its properties and behaviors that must be taken into the consideration for cost feature implementation and applications. Figure 5.4 illustrates the partial relations among different features for cost engineering, in which the process cost feature and product batch cost features are presented.

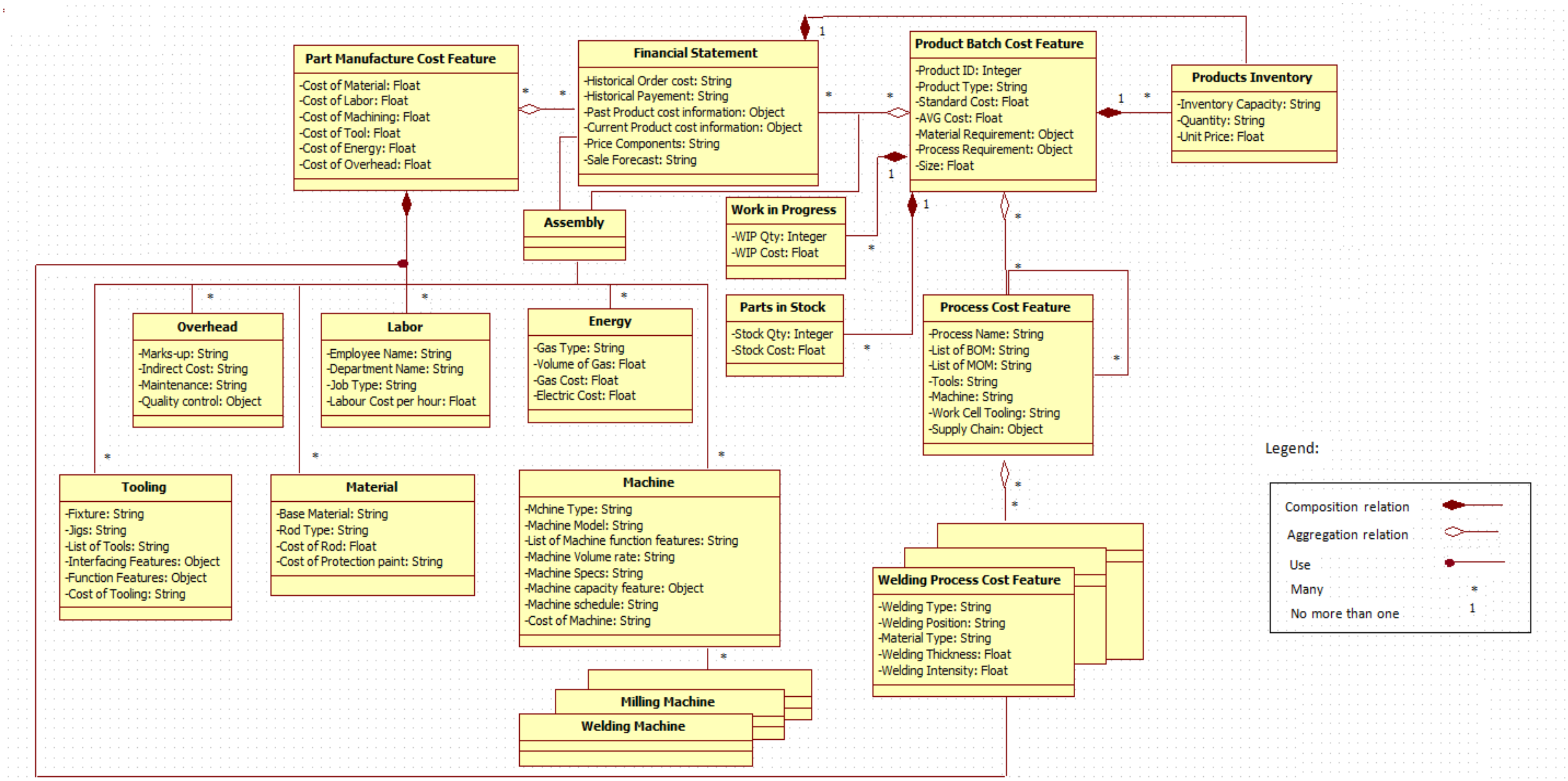


Figure 5.4 Cost feature definition in UML format

For extracting the cost feature, we need to provide reference data that includes specific design and manufacturing information, such as the number of available machines in the shop floor, the capacity of each machine, the operation that can be performed by each machine, the available human resources that can work with each machine, and the sequence of the machining process. Then, some functions are required to extract the cost features and related information from the design feature model, manufacturing process features, and ERP system.

According to [43,46] for a welding process cost feature, which can be a child class of the process cost feature, as shown in figure 5.4, the following welding feature class is suggested:

Where constraints means:

- L_{ij} (): welding length

- SS (): welding section

- ρ (): the rod output

- d (): the filling metal density

- η (): the process efficiency

- D_a (): protection gas-consuming time

- φ (): operator efficiency

- p (): degree of weld position complexity

- M_h (): unit time cost ($/hour) for machine h

- S_h (): setup cost ($/hour) for machine h

Where functions means:

- V_s (): welding volume evaluation: $V_s = SS \: x \: L_{ij}$

- m_a (): amount of welding wire: $m_a = \frac{\rho \text{ x } V_s \text{ x } d}{\eta}$

- t_{arc} (): electrode time: $t_{arc} = \frac{m_a}{D_a}$

- T_{ij} (): welding operating time: $T_{ij} = p \text{ x} \frac{t_{arc}}{\varphi}$

- C_{ij} (): machining cost for welding: $C_{ij} = M_h \text{ x } T_{ij} + S_h$

Table 5.1 Cost feature template

Feature ID	Feature type	Process step	Process type	Process description	Welding sections	Welding operating time	Machining cost	Material cost
xx-01	T-joint	1	Shielded metal arc welding	End to end with vertical borders	(e, $\alpha, g, t, h, form$)	T_{ij}	C_{ij}	M_{ij}

Regarding extracting information from parts' geometry and process features, some of the weld information has to be inquired from the CAD model so that the welding time, welding length, volume of welding wire (rod), and the welding tool as well as its control can be worked out.

As table 5.1 illustrates the cost component variables identified in the cost features are to be used as the initial input for both the empirical estimation as well as the data mining (DM) analysis. Although it is ideal to achieve the above-mentioned data extraction by automatic algorithms embedded in CAD/CAM tools, engineers' interactive input through friendly and consistent user interfaces can also be accepted. Hence, the application barrier for this proposed cost-estimation method can be significantly lowered.

5.5 Empirical cost estimation process

Usually, companies use the parametric cost calculation formula to determine the cost of each product. However, they are not considering all of the products' features when they use this formula. As a result, their cost calculations are not accurate. To solve this problem, we used empirical analysis to become confident that we considered all of the products' features to achieve an accurate CE.

Figure 5.6 shows the proposed empirical CE process built into the CE engine. The empirical CE module uses the existing ERP data and interpolates data to analyze the weights and coefficients of cost component initial variables from the cost feature input. Regression analysis is a useful tool for evaluating the weights of variables and the coefficients among them. Regression curves can show the data strength and their relations. By using the above analysis, the relative statistic cost weight predictions and coefficients can be extracted out of the initial variable set; afterward, those contributing and independent variables are quantitatively used to form a linear regression CE formula.

A cost data set is defined as $\{C_{total}, x_1, x_2, \dots, x_n\}$ where the total cost of the product is denoted by C_{total}; it is a differentiable function of n input component variables: $x_1, x_2, \dots, x_n$.

The goal of regression is to draw a line through the data set that can pass data as much as possible. To find the best-fit line or curve of the data, the regression has to analyze the relationship between the variables that are denoted by w_i to illustrate the coefficient among x_i and C_{total}, which means how changes in x_i can affect C_{total}.

$$C_{total} = w_1x_1 + w_2x_2, \dots + w_nx_n + \varepsilon \qquad (5\text{-}1)$$

We need to find the best set of $\{ w_i \}$ to minimize the error ε and build a more accurate formula for CE. To achieve this goal, the data related to each variable has to be analyzed to qualify the product data with high precision. For example, observing the amount of welding wire consumed in 40 products and their relation to total cost is shown in figure 5.5. After analyzing the data of each variable, we can use its coefficient to build the CE formula.

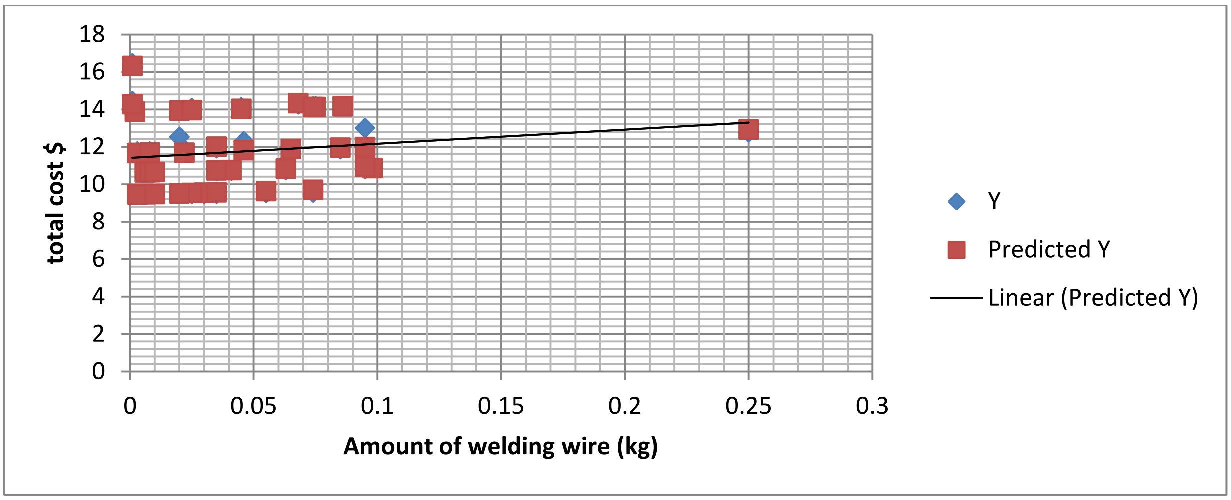


Figure 5.5 Relation between total cost and amount of welding wire

The CE formula needs to be updated after the evaluation with the existing pool of the existing data, and then, it can predict the initial cost and its accuracy for a new product or order. Along the way, the estimation confidence level is constantly checked. If the predicted costs cannot meet the confidence level required, the weights of the regression analysis have to be updated with regression refinement. This process will be iterated several times to ensure that the CE formula and the level of accuracy are acceptable.

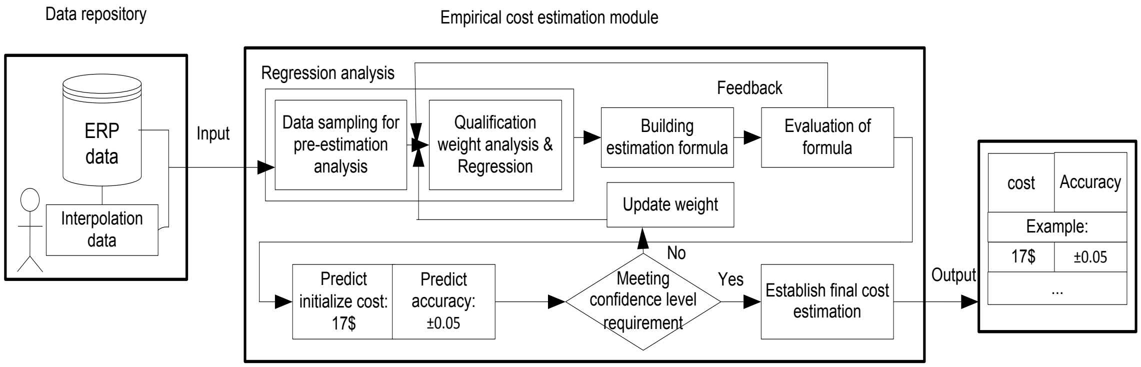


Figure 5.6 Empirical cost estimation process

Five hundred products were selected as sample data elements for testing regression analysis for the purpose of CE. The output of regression analysis between the total cost C_{total} and $\{ X_i \}$ shows that Multiple R (correlation coefficient) is 0.99, which is high. The standard range of Multiple R is between -1 and 1, in which 1 is showing a strong relation among the variables. In addition, R Square (coefficient of determination) is 0.97, which is also high. The range of R Square has to be between 0 and 1, and 1 means that the regression line passes exactly through the data. Figure 5.7 shows the scatter plot of the multiple regressions that have been generated by regression analysis. Note that the products have been organized according to the individual total costs.

Table 5.2 Regression Statistics Result

Multiple R	0.99
R Square	0.97
Adjusted R Square	0.85
Standard Error	0.33
Observations	500

The formula extract from regression:

Welding cost = 1.2812E-13 + (Labor cost per hour × -0.059562) + (Amount of welding wire ×1.44) + (paint cost×1) + Gas volume rate × 10.5) + (Machine cost × 4.2) + (overhead cost × -0.621) (5-2)

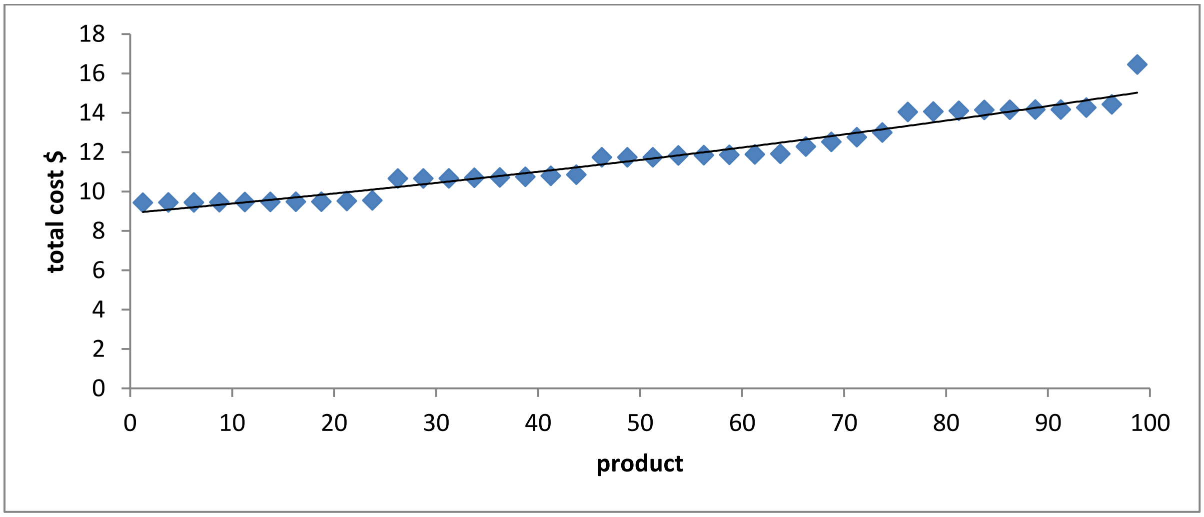


Figure 5.7 Sample of scatter plot of the multiple regression analysis

5.6 DM cost estimation process

An accurate CE can be seen as an optimization problem in which we have to search for the most accurate estimate. Data mining is a powerful technique that can be used in CE. Indeed, there are many data mining algorithms, and each is suitable for one kind of data modeling. Some widely used techniques such as linear regression can describe data coefficients and behavior.

Data mining can extract knowledge from a significant amount of data, transfer the new data into useful information, and discover useful information and hidden patterns by analyzing the data. In the recent decade, data mining has been widely used for data training and testing in order to extract data patterns.

The proposed DM CE method has several interfaces with the product domain and the process domain. Figure 5.8 illustrates the DM cost estimation configuration. The product domain contains information about design configurations, modules, parts, design features, and materials. The process domain defines manufacturing processes, such as machining operations and their related machines, tools, and labors involved.

On the right side of figure 5.8, the DM cost estimation module is shown. It can calculate the cost of each specific manufacturing feature based on the information of the product and process domains as well as the ERP historical data. To achieve this goal, the proposed DM module uses ERP historical data to categorize cost data by product families based on product configurations, materials, and manufacturing processes. In the next step, the DM module analyzes the cost data to associate the cost values with cost feature variables. Then an appropriate data mining algorithm is selected and used to extract the full cost feature patterns (i.e. object instances). As a result, the selected data mining algorithm is associated with the sample data product family. In implementation, the user interface will show the confidence level of each algorithm to the user, and based on that, the user can select the best algorithm for each product family (i.e. group).

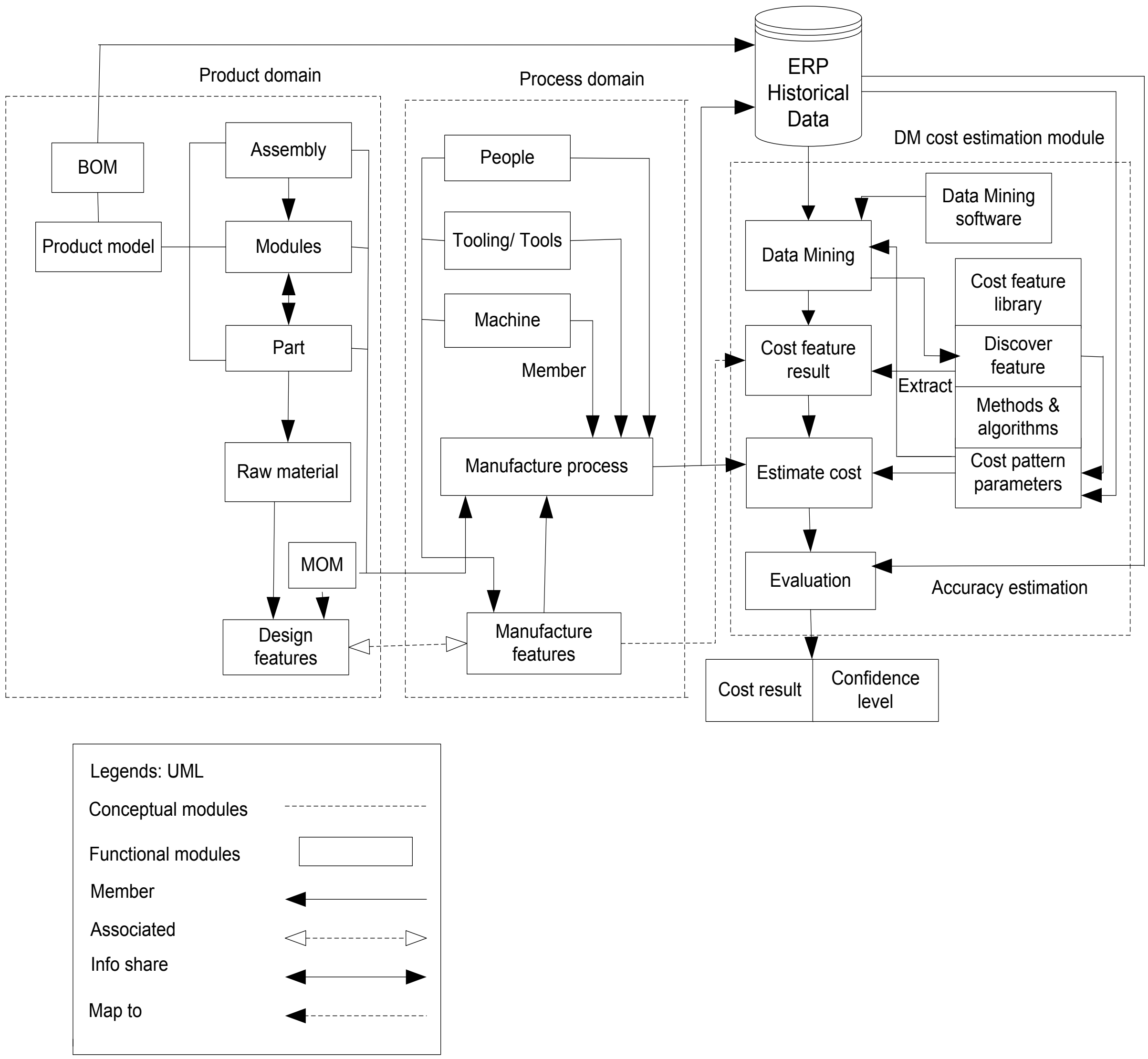


Figure 5.8 DM cost estimation configuration

5.7 Data mining implementation in the cost engine

For implementing the proposed DM cost estimation method, the cost-related data has been saved in the database. Several data mining algorithms have been implemented and tested in WEKA to extract and process this significant amount of data in an efficient manner. WEKA has been selected in this research because it is machine learning software that provides a comprehensive collection of machine learning algorithms for automatic classification, regression, clustering, and feature selection [79].

Usually, SCEs use different algorithms and Artificial Intelligent (AI) techniques. Such as Linear regressions, multiple regressions [80], and artificial neural networks (ANNs) [81, 82] which are commonly applied in many types of CE software. They reported that these algorithms are more suitable for numerical data, and ANN presents more accurate data with a lower error percentage [80]. Besides the previously mentioned techniques, we added KNN and MP5 data mining algorithms, which are suitable for both numerical and non-numerical data and multiple relations [83].

In the next step, the results of the DM algorithms were used to generate new rules and patterns for welding CE. Then, the result of the algorithm was saved as a model to apply to cost estimation. Further, the user has to select the most effective CE algorithm based on confidence level. Confidence level can be defined by comparing actual effort and estimated effort based on the mean magnitude of relative error (MMRE). MMRE is a standard function used in SCE that is applied to evaluate the performance of CE on a training data set [53]. Where n denotes the number of manufacturing parts used for CE and MRE is equal to the following:

$$\text{MMRE} = \frac{1}{n} * \sum_{i-1}^{n} MRE \qquad (5\text{-}3)$$

$$\text{MRE} = \frac{|\text{Actual Cost} - \text{Estimated Cost}|}{\text{Actual Cost}} \qquad (5\text{-}4)$$

Chapter 6: Case study – welding process cost estimation

6.1 Definition of welding process features

Welding features are a new type of manufacturing features defined in this work, which have not been formally defined in the literature thus far. Figure 6.1 shows a typical welding feature that currently uses. In this work, the welding feature is a child class of the *manufacture feature* (see Figure 6.2) , and it contains a set of extracted welding related process data, which is associated with welding geometric entities and their characteristic parameters, technical process setup parameters, and related resources. Welding-related geometric entities are further related to product design features and manufacturing derivatives, such as welding paths and which angles have to be defined within any welding feature.

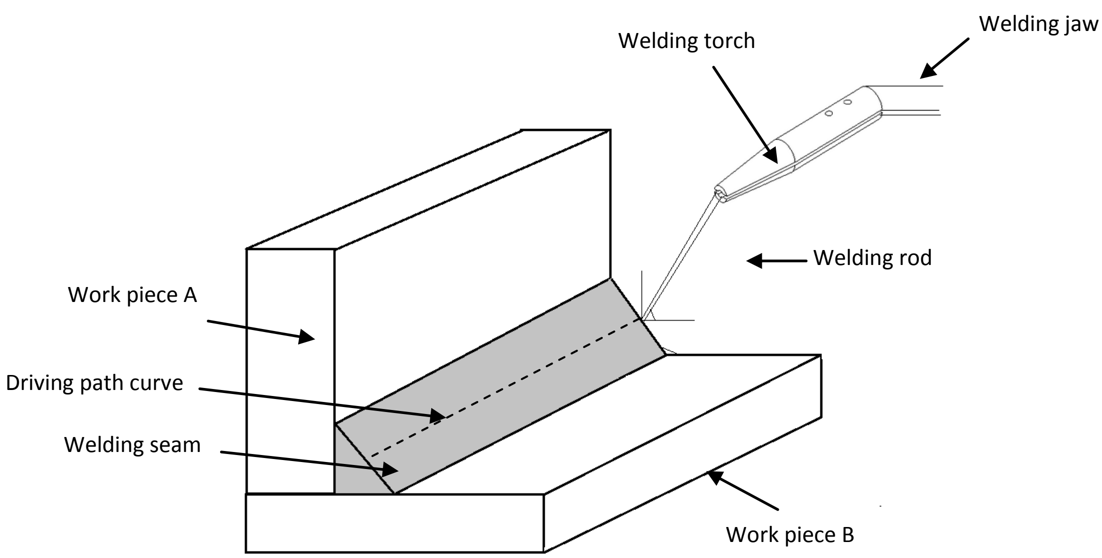


Figure 6.1 A typical welding feature

Referring to the welding feature class as shown in figure 6.2, the welding feature operation process is executed via three main predefined class methods: *preparing-welding(), run-welding(), and after-welding().* They interact internally with the feature properties as shown in the figure. For example, the preparing process is defined with the setup information; it extracts technical parameters from the design features, related manufacturing features, and generates the necessary drive curves and path segments as well as characteristic geometric parameters from product design and structure entities. Note there is a unified product feature model shown in figure 6.2. The welding process planning function uses the manufacturing features from the product unified feature model as well. The *run-welding()* methods uses the real-time information of the assigned welding equipment resources from workshop process planning and operation model via their data object pointers, such as those describing welding machine and welding tools. Finally, *after-welding()* methods is supposed to define and guide operational application to do cleaning, inspection, and test tasks.

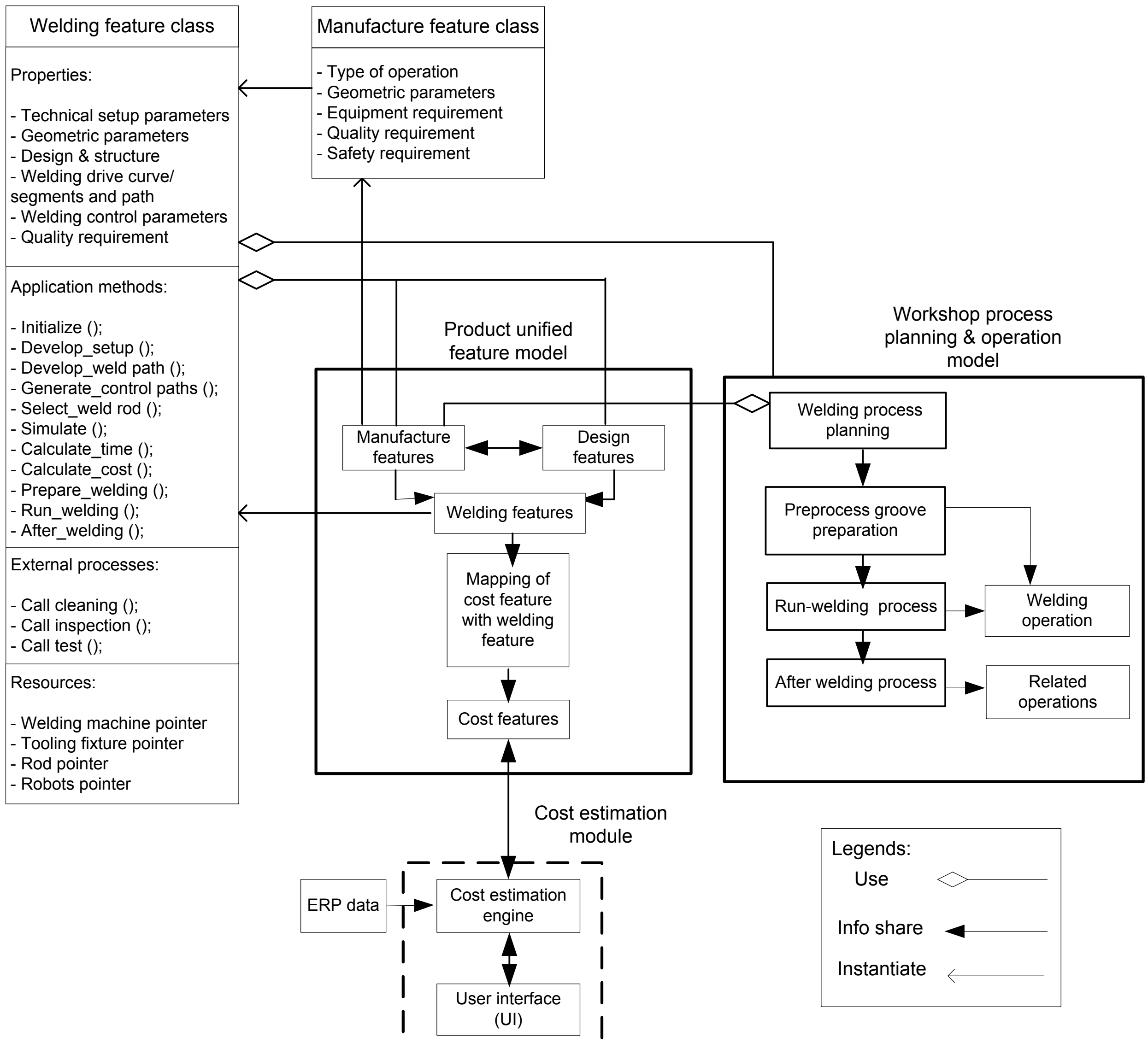


Figure 6.2 Welding feature and cost feature structure relations

This case study has implemented a preliminary CE engine that works with welding features whose template is defined as a feature class. In addition to the live support from the instances of functional modules as indicated in figure 5.8, the feature library will collect all the information related to welding features to classify, sort, and analyze them for future data mining processes.

6.2 The data structure of welding feature

In this research, the following data structure of welding feature class is suggested:

- L_1, W_1, H_1: length, width, and height of first piece
- L_2, W_2, H_2: length, width, and height of second piece
- T_1: time of weld
- T_2: time of setup
- T_3: time of paint
- T_4: time of clean
- BM: Base Material
- WThick: weld thickness
- WType: weld type
- Tech: technology used in welding
- WPosition: welding position
- WProcess: welding process
- WW: amount of welding wire
- PP: amount of protection paint
- GT: gas type
- GV: gas volume rate
- OF: operating factors
- StartP: the location of start point
- EndP: the location of end point

Data structure of welding feature class
Properties: - Geometry information of first work piece: L_1, W_1, H_1 - Geometry information of second work piece: L_2, W_2, H_2 - Welding times: T_1, T_2, T_3, T_4 - Technical parameters - Welding geometric parameters ……………
Constraints: - $L_1 > 5$cm - $90° < \alpha < 180°$ - WThick < 10 cm …………….
Functions: - Get-parameters (); - Save-parameters (); - Calculate-welding-time (); - Initiate-feature (); - Initiate-UI (); ……………….

Figure 6.3 Data structure of Welding feature class

- form: form has three positions: End to End in V; End to End with vertical borders; and End to End in X. Figure 6.4 illustrates these three positions:

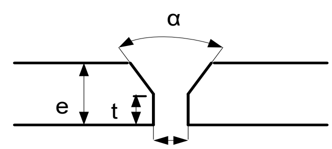
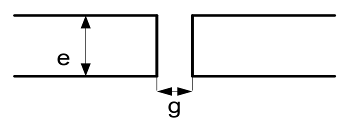
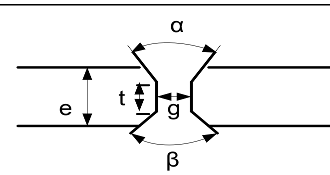
Weld feature objects	**Diagram**
End to End in V - e: height of thickness - t: heel or height of the flat - α: chamfer opening top angle	
End to End with vertical borders - e: height of thickness - g: clearance apace	
End to End in X - e: height of thickness - α: chamfer opening top angle - β: chamfer opening bottom angle - g: clearance apace - t: heel or height of the flat	

Figure 6.4 Geometric Parameters of welding in different welding form [46]

According to the current definition, P1 represents the geometric information of the first piece such as length, width, and height. P2 represents the geometric information of the second piece that needs to be welded to the first piece. TP includes the technical parameters of welding, which includes base material, weld thickness, weld type, technology used in welding, welding position, welding process, amount of welding wire, amount of protection paint, gas type, gas volume rate, operating factors, the location of start point and end points of welding. GP is the geometric parameters of welding, which includes shape characteristics. Table 6.1 illustrates the data structure of the welding feature properties:

Table 6.1 Data structure of welding feature properties

welding feature properties	abbreviation	Unit
Base Material	BM	Inch, SQIN
Weld Thickness	WThick	mm
Weld Type	WType	Grove, Seam, Joint
Technique of Grove	Tech Grove	U, square, flare, bevel, V
Technique of Seam	Tech Seam	single side, double side, all round
Technique of Joint	Tech joint	Square butt joint, V butt joint, Lap joint, T-joint
Weld Position	WPosition	axis(x, y, z)
Weld Process	Wprocess	SMAW, SAW, GMAW, GTAW
Gas type	Gas type	CO2 , argon, argon/co2 mix
Operation factor	OF	(SMAW, 30%), (GMAW, 45%), (MCAW, 55%), (GMAW, FCAW, and SAW processes, 100%)
Start point	StartP	(x,y,z)
End point	EndP	(x,y,z)

6.3 The data structure of welding cost feature

By using the data structure of welding, we can define welding cost feature class as:

- Cost of material (C mtl) = (amount of welding wire× C Rod) + (amount of protection paint × C Protection) + (gas volume rate × C Gas)

- Cost of labor (C lab) = (t) × C labor × Operating Factor

- Cost of machine (C mch) = (C occp × T_1) + (C setup × T_2) + (C clean × T_3) + (C paint × T_4)

- Cost of welding (C weld) = f (Type of welding, welding condition)

- Cost of overhead (C OH) = f(C administration, C depreciation, Infrastructure)

CF (): cost feature: CF = $[C_{mtl}, C_{lab}, C_{mch}, C_{weld}, C_{overhead}]$

Data structure of welding cost feature class
Properties: - Geometry information of first work piece: L_1, W_1, H_1 - Geometry information of second work piece: L_2, W_2, H_2 - Welding times: T_1, T_2, T_3, T_4 - Technical parameters - Welding geometric parameters ……………
Constraints: - C_{lab} = \$24 per hr - $C_{mch} < T_1 * 3\$$ ……………
Functions: - Get-parameters (); - Calculate-cost of material (); - Calculate-cost of labor (); - Calculate-cost of machine (); ……………….

Figure 6.5 Data structure of welding cost feature class

6.4 Mapping of welding manufacture feature and the associative cost feature

To illustrate the cost feature association method, a feature-mapping system was used. Feature-mapping system can map the cost features with other types of features in CE via a set of data structures, and this data was derived from a CAD module or an ERP system. The mapping function defined as a: MF = [WF, CF, f] where: f (WF) = CF

WF means welding feature, CF is the cost feature, and f represents the function mapping algorithm. In the matrix below, we present a feature-mapping schematic between the welding and cost features.

Mapping matrix Feature array Cost feature

$$\begin{bmatrix} C_{Rod} & C_{Pro} & C_{Gas} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{Lab} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{Occp} & C_{Set} & C_{Clean} & C_{Paint} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C_{Weld} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C_{OH} \end{bmatrix} * \begin{bmatrix} WW \\ PP \\ GV \\ t * OF \\ T_{weld} \\ T_{set} \\ T_{clean} \\ T_{paint} \\ f_{weld} \\ f_{OH} \end{bmatrix} = \begin{bmatrix} C_{mtl} \\ C_{lab} \\ C_{mch} \\ C_{weld} \\ C_{overhead} \end{bmatrix}$$

(6-1)

6.5 Prototype results and discussion

To test the advantages of the proposed hybrid feature-based CE, a typical problem was formulated as discussed in chapter 5 and the computer program was implemented. Visual Studio 2012 was used to implement the CE engine as a functional module acting as a driver in the prototype program; and an SQL Server 2012 was used to store and manage our features and ERP data.

To obtain preliminary results, we selected a sample set of welding-related operation data with 500 welding parts, currently manufactured by McCoy FARR, ranging from simplistic to complex. This test case is large enough to serve as a good benchmark for our proposed formulation, and yet small enough to be tractable, thus allowing us to test our hypothesis. The test case was selected to mimic real welded parts and a real welding feature-based CE problem as close as possible to reality, ensuring the validity of the predicted results.

In the first step, we extracted data from the company's ERP system for data analysis. Recognizing the trend data, reasonably established by common numerical analysis, is the next step. Data trends can illustrate the overall pattern of data changes, which is suitable for comparing different groups of data. To determine data trends, more details need to be added, such as machining and overhead costs. Besides adding more details, the new welding cost is calculated based on a regression cost formula, which was introduced in section 5.6. The data analysis shows that the new cost increases 7%-9% compared to the primary company's data.

The regression welding CE is presented below in table 6.2, beside the primary welding cost. As a result, accurate CE based on a forward approach is used to cross validate the DM cost estimation process.

Table 6.2 Sample CE based on regression analysis

Part ID	Joint type	Labor cost per hour ($)	Welding time (hr.)	Operating factor (%)	Amount of welding wire (kg)	Per rod cost ($)	Paint cost ($)	Gas volume rate (m^3/hr)	Gas cost per cubic meter ($)	Machine cost ($)	overhead cost ($)	Primary weld cost ($)	Regression weld cost ($)	Predicted cost ($)
A101	butt-joint	24	0.16	0.4	0.03	1.44	0.25	0.62	10.5	0.48	0.65	8.43	9.438	9.427
A102	butt-joint	24	0.25	0.4	0.075	1.44	0.25	0.62	10.5	0.75	0.8	9.65	10.859	11.17
A103	butt-joint	24	0.33	0.4	0.062	1.44	0.25	0.62	10.5	0.99	0.9	10.5	11.85	11.849
A104	butt-joint	24	0.5	0.4	0.75	1.44	0.25	0.62	10.5	1.5	0.98	12.62	14.156	14.155
A105	T-joint	24	0.16	0.4	0.06	1.44	0.25	0.62	10.5	0.48	0.66	8.44	9.449	9.449
A106	butt-joint	24	0.33	0.4	0.03	1.44	0.25	0.62	10.5	0.99	0.75	10.47	11.742	11.742
A107	butt-joint	24	0.16	0.4	0.25	1.44	0.25	0.62	10.5	0.48	0.82	11.1	12.763	12.52
A108	Lap-joint	24	0.33	0.4	0.046	1.44	0.5	0.84	10.5	0.99	0.76	10.54	12.288	11.801

Based on existing information for welding costs and features, we need to design a set of rules (tailored to a specific implementation) that allowed us to define the guidelines used to implement the DM cost estimation technique consistently, for the case study. The set of rules we defined is the following:

- Rule 1: Joint type: {butt-joint, Lap-joint, T-joint}
- Rule 2: Labor cost ($): {hourly = 24}
- Rule 3: Welding time: {setup time, length of weld, welding sections, complexity, machining volume rate}
- Rule 4: Operating factors: {FCAW= 40%}
- Rule 5: Amount of welding wire: {welding volume, filling metal density}
- Rule 6: Rod cost($): {per rod 7018=1.44}
- Rule 7: Paint cost($): {welding sections, joint type (butt-joint = 0.25, Lap-joint = 0. 5, T-joint = 0.25)}
- Rule 8: Gas volume rate: {nozzle size ½ inch = 0.62 (m^3/hr.), nozzle size 5/8 inch = 0.84(m^3/hr.)}
- Rule 9: Gas cost($): {per cubic meter= 10 }
- Rule 10: Machine cost($): {welding time *3}
- Rule 11: Overhead cost($): {0.05*direct cost}

The proposed feature-based cost engine uses the following steps to prepare DM cost estimation:

1. Reorganizing and setting cost feature data extracted from the ERP system
2. Implement data mining clustering algorithm via WEKA: Expectation Maximization (EM) selected as a data mining clustering algorithm, which models a data set based on linear combination and normal distribution [84]. EM divided a data set into seven clusters, and then each cluster was used as an input for CE. Eighty percent of each cluster was used as training data set and 20% for testing the data set.
3. Implement data mining algorithms via WEKA: Linear Regression (LR); Multiple Regression (MR); Artificial Neural Network (ANN), k-Nearest Neighbors

(KNN), and the Meteorological parameter section (MP5). For instance, figure 6.6 illustrate the implementation of the MP5 algorithm in WEKA. This model, generated for each cluster, was saved and reused for further CE.

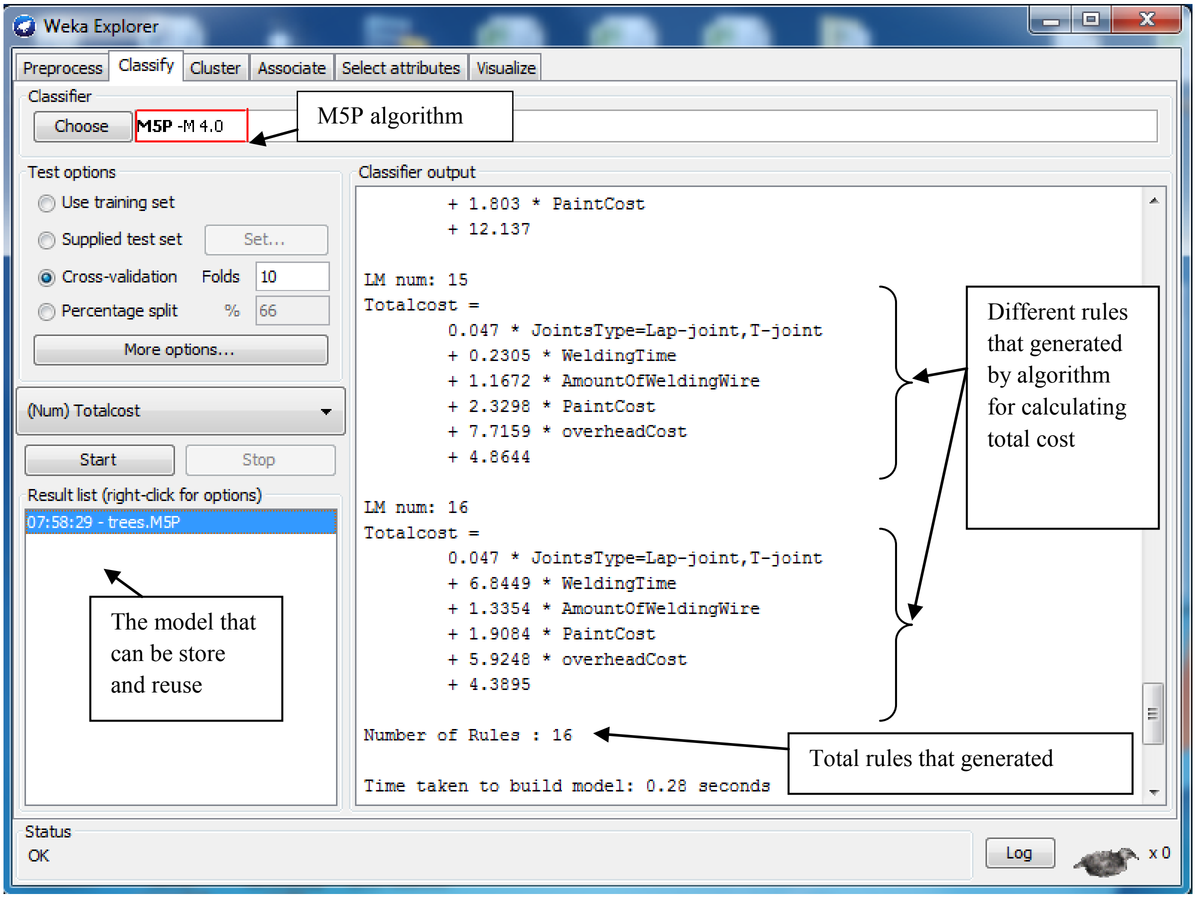


Figure 6.6 Sample result of CE based on the regression analysis of weka

Five hundred welding parts were used as an input for five generated data mining models. The comparison between the results of five different algorithms in the case study is shown in table 6.3 Clearly, ANN and MP5 have better results for correct prediction. In addition, ANN has a better confidence level compared to the other algorithms. To optimize the result of models, the trends of MMRE were analyzed. LR can be optimized by adding MMRE to the CE. However, other models do not have a same trend, so the result is sometimes less than the actual cost and sometimes more. In this research, the suggested algorithm for welding CE is the artificial neural networks (ANN) algorithm.

Table 6.3 correct matching and error rates comparison between five different algorithms to the historical data set (%)

Models	N	LR			MR			ANN			KNN			MP5		
Correct prediction		training data	testing data	MMRE	training data	testing data	MMRE	training data	testing data	MMRE	training data	testing data	MMRE	training data	testing data	MMRE
Cluster 0	52	78%	53%	53.4%	74%	68%	33%	84%	92%	16.8%	72%	86%	51.6%	85%	95%	33%
Cluster 1	130	68%	43%	52.1%	69.3%	29.6%	52%	82.3%	96.2%	15.6%	69.3%	35%	65.6%	79.6%	92.3%	25.3%
Cluster 2	78	82%	85%	45%	85%	89.6%	16.2%	79.6%	93.6%	16.5%	86%	70.3%	39.6%	74.6%	89.6%	35.6%
Cluster 3	43	76%	69%	33%	63.3%	46.2%	39.3%	85.6%	85%	12.3%	64.3%	35.7%	52.3%	89.7%	92.6%	55.6%
Cluster 4	19	68%	73%	52.3%	69.6%	36.1%	23.6%	80.3%	86.3%	16.3%	82.6%	73.5%	33.5%	86%	92.3%	38.9%
Cluster 5	55	64%	35%	59.6%	67.4%	42.5%	17.6%	78.6%	92%	19.3%	68.9%	75%	59%	84%	85.6%	18.9%
Cluster 6	123	74%	46%	43.8%	72.6%	39.6%	56.4%	83.6%	93.5%	25.2%	74.8%	48.6%	45%	84.6%	91.6%	47.6%
AVG		73%	58%	48%	72%	50%	34%	82%	91%	17%	74%	61%	50%	83%	91%	36%

Figure 6.7 illustrates more extended welding cost information, which includes following inputs:

-Joint type: butt-joint, Lap-joint, T-joint

-Welding length: the length of welding seam (mm)

-Welding time: setup time, paint time, clean time, weld pointing time, total welding time (hr)

-Operating factor: (%)

-Labor cost: labor cost per hour, labor working cost ($)

-Paint cost: ($)

-Rod cost: amount of welding wire, rod diameter, per rod cost, total rod cost ($)

-Power cost: ($)

-Machine cost: ($)

-Overhead cost: ($)

-Gas cost: gas volume rate, amount of gas used, gas cost per cubic meter, type of gas, gas cost by feature, total gas cost ($)

This kind of information should be provide by user, however if some data is not available, the UI will be fill it based on historical data. This UI is intended to allow cost engineers to review and capture all welding cost features; then the cost can be calculated based on parametric CE and five data mining algorithms. Finally, the company can select the best algorithm based on the company’s strategy for CE.

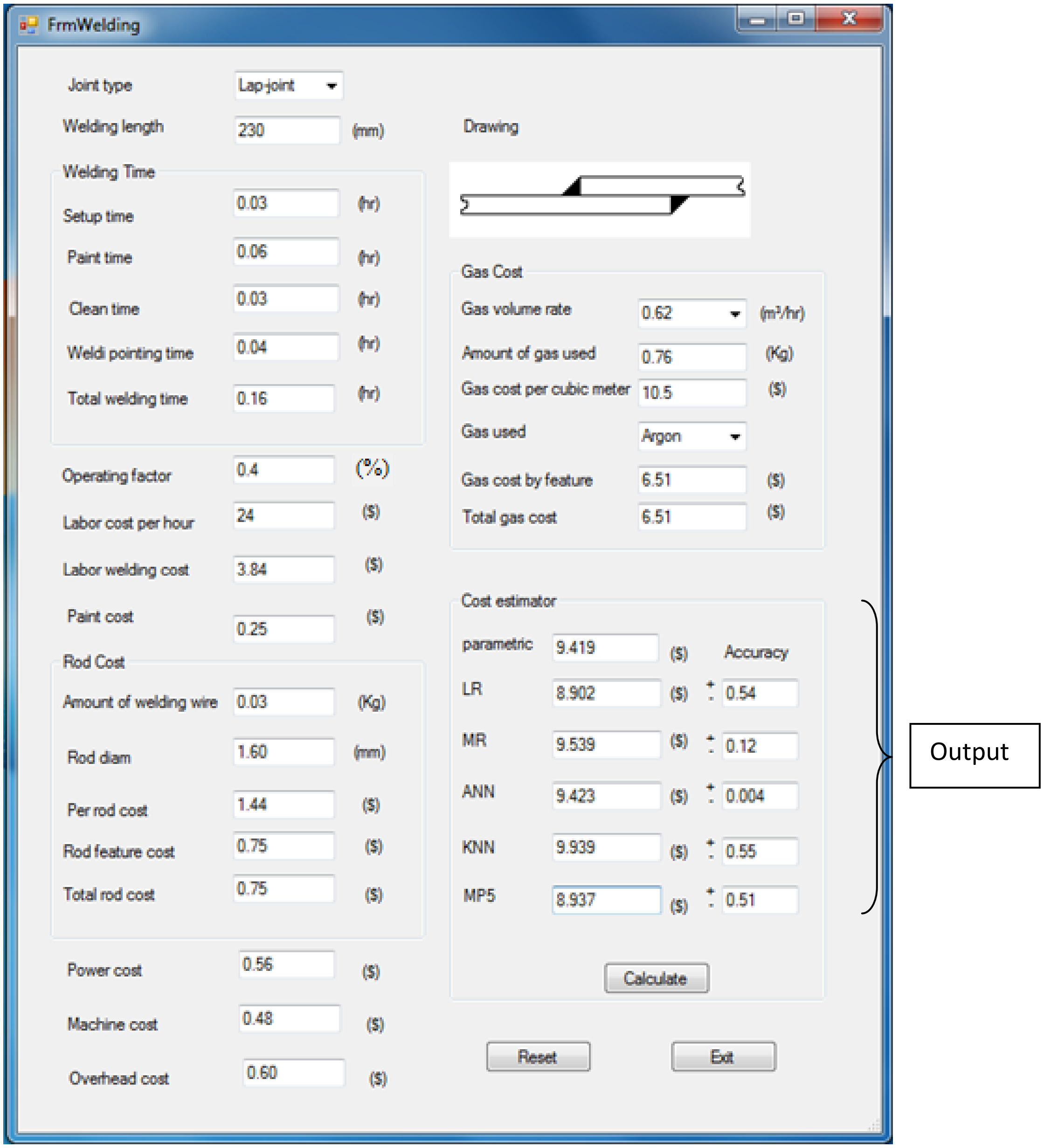


Figure 6.7 User Interface (UI) for welding cost estimation

6.6 Conclusion

This research presents a hybrid CE methodology with feature-based empirical data regression and data mining algorithms. The case study shows an application to welding products, and it demonstrated that the approach is capable of producing satisfactory results for different welding-feature-based parts. The proposed model has proven to be repeatable, robust, and accurate for the range of features applied. In addition, as shown in the case study, a set of rules or guidelines used to fill the gaps in missing data. Such rules are complimentary with specific applications while the built-in features consider general patterns of welding parameter estimation. This estimation methodology could be extended beyond the cost of engineering only, e.g., the total manufacturing time evaluation.

However, there are some limitations of which the reader should be aware:

- Predefinition of manufacturing features is the first limitation. The manufacture process data can be consistently managed and analyzed. The consistent definitions need one-time semantic modeling effort in any adopting enterprise. However, reusing the definitions with repetitive application is highly recommended and hence the implementation barrier would be reduced significantly.
- Data mining has to be supported with available and comprehensive historical data.
- The another limitation of this proposed method is that currently, in our case, ERP data has to translated into a different data structure in order to support data mining and feature data extraction. However, hopefully, with the popular adoption of the proposed scheme and well-defined feature. Then, the ERP data structure can be interfaced automatically with a converter or a compliance standard. Clearly, a lot of future work is necessary.
- The proposed system is based on a semi-automatic hybrid method, where categorizing the product features based on their shapes and characteristics has to

be done by engineers. Potentially, this task could be assisted by a feature recognition program [85], which is not the focus of this study.

- The company's ERP data description is not the same as our welding and cost feature- definitions. We preprocessed the data based on the definitions proposed, to keep the consistency of data input to the prototype algorithm. This issue arose from the fact that in some instances, the actual manufacturing method applied is not the same as the process plan, causing some discrepancy in our testing model.
- Finally, some features values do not exist in ERP system and we have to find them based on other sources. For example, to find the thickness of welding specifications of the seam, information has to be extracted from a CAD/CAM model.

Chapter 7: Conclusion and future work

7.1 Conclusion

The main subject of this research was developing a concept of unique feature type for cost estimation purpose. A new category of feature, the *cost feature* is defined as a class in the unified feature modelling system to address the characteristics of cost engineering entities, constraints and dependency relations.

Firstly, this study proposes the semantic model for automatic product cost estimation. This model integrates three functional sub-models: feature-based costing, data mining, and semantic reasoning. The goal of this model is investigate a concept of new manufacturing cost calculation model coherently throughout the lifecycle of a product series, especially emphasizing at the conceptual design stage.

Secondly, this study describes a new cost estimation functional module that can be implemented in an ERP system including the representative of auxiliary data, machining and design models. Some key concepts in the module, such as cost feature library, the relation among sub-models and feature mapping are proposed.

Thirdly, this study proposes a hybrid cost estimation method by combining traditional empirical and advanced data-mining techniques. The goal of this method is providing the more accurate cost estimation via contrasting a user interface that can show the cost estimation cross-checking.

The author believes that there are, significant advantages in implementing the proposed hybrid methodology:

- Capability to work out the estimated confidence level from different components
- Less time and complexity in application compared to the traditional analytical cost-estimation formula
- Reusability and recursive accuracy enhancement over cost patterns and estimations from DM results based on accumulated historical data
- Support new part CE without detailed exhaustive data
- Feasibility of applying in a manufacturing company within a ERP system via customized UIs for increased CE accuracy and short feedback time to the decision-makers and customers
- A case that may benefit from the methods proposed in this case study
- The results of case study support this argument and provide additional guidelines to cost engineers

7.2 Future work

This study has illustrates the benefits of using cost feature for dynamic, complex, and bottom-up cost estimation. Furthermore, the results suggested the more accurate data mining algorithm for welding cost estimation. However, it is to be further proven whether other companies will experience similar improvements. To address the generality of applications, future work must be done in the following fields:

- Develop a comprehensive cost estimation model for other types of operations rather than welding
- Advanced cost analysis for products with two or more operations
- Apply method to other industries

References

[1] Kaplan, Robert S., ed. Measures for manufacturing excellence. Harvard Business Press, 1990.

[2] Mileham, A. R., G. C. Currie, A. W. Miles, and D. T. Bradford. "A parametric approach to cost estimating at the conceptual stage of design." Journal of engineering design 4, no. 2 (1993): 117-125.

[3] Fritz, Ronald D., and D. L. Kimbler. "Traditional product cost estimation: problems and alternative"." In Proceedings of the 5th Industrial Engineering Research Conference, Norcross, GA, pp. 375-80. 1996.

[4] Robert Jacobs, F. "Enterprise resource planning (ERP)—A brief history."Journal of Operations Management 25, no. 2 (2007): 357-363.

[5] Daneva, Maya, and Roel Wieringa. "Cost estimation for cross-organizational ERP projects: research perspectives." Software Quality Journal 16, no. 3 (2008): 459-481.

[6] Hong, Kyung-Kwon, and Young-Gul Kim. "The critical success factors for ERP implementation: an organizational fit perspective." Information & Management40, no. 1 (2002): 25-40.

[7] Herrmann, Jeffrey W. "Design for production: concepts and applications." In Proceedings of the SME East Coast Region 3 Annual Members Conference, Bethlehem, PA, USA, October 10-11. Hiller Aircraft Corporation, 2003.

[8] Woods, Donald R. Cost Estimation for the Process Industries. McMaster University, 1983.

[9] Trivailo, O., M. Sippel, and Y. A. Şekercioğlu. "Review of hardware cost estimation methods, models and tools applied to early phases of space mission planning." Progress in Aerospace Sciences 53 (2012): 1-17.

[10] Holmes, Geoffrey, Andrew Donkin, and Ian H. Witten. "Weka: A machine learning workbench." Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on. IEEE, 1994.

[11] Ma, Yongsheng. Semantic Modeling and Interoperability in Product and Process Engineering. Springer London Limited, 2013.

[12] Niazi, Adnan, Jian S. Dai, Stavroula Balabani, and Lakmal Seneviratne. "Product cost estimation: Technique classification and methodology review." Journal of manufacturing science and engineering 128, no. 2 (2006): 563-575.

[13] Clark, Forrest, and Albert B. Lorenzoni. Applied cost engineering. CRC Press, 2011.

[14] Stewart, Rodney D., Richard M. Wyskida, and James D. Johannes, eds. Cost estimator's reference manual. Vol. 15. John Wiley & Sons, 1995.

[15] Rush, Christopher, and Rajkumar Roy. "Analysis of cost estimating processes used within a concurrent engineering environment throughout a product life cycle." 7th ISPE International Conference on Concurrent Engineering: Research and Applications, Lyon, France, July 17th-20th, Technomic Inc., Pennsylvania USA. 2000.

[16] Ansari, Shahid L., et al. Target costing. Irwin/McGraw-Hill, 1997.

[17] Roy, Rajkumar. "Cost engineering: why, what and how?." (2003).

[18] Ostwald, Phillip F., and Timothy S. McLaren. Cost analysis and estimating for engineering and management. Pearson Education, 2004.

[19] Sweeting, Jack. Project Cost Estimating: Principles and Practice. IChemE, 1997.

[20] Shuford Jr, R. H. "Activity-based costing and traditional cost allocation structures." Cost estimator’s reference manual, 2nd ed., New York, Johan Wiley & Sons Inc (1995): 41-94.

[21] Brinke, Erik. Costing support and cost control in manufacturing. A cost estimation tool applied in the sheet metal domain. University of Twente, 2002.

[22] Asiedu, Y., and P. Gu. "Product life cycle cost analysis: state of the art review." International journal of production research 36.4 (1998): 883-908.

[23] Hart, C. G., et al. "An advanced cost estimation methodology for engineering systems." Systems Engineering 15.1 (2012): 28-40.

[24] Roy, Rajkumar, et al. "Quantitative and qualitative cost estimating for engineering design." Journal of Engineering Design 12.2 (2001): 147-162.

[25] Layer, Alexander, et al. "Recent and future trends in cost estimation."International Journal of Computer Integrated Manufacturing 15.6 (2002): 499-510.

[26] Datta, Partha P., and Rajkumar Roy. "Cost modelling techniques for availability type service support contracts: a literature review and empirical study." CIRP Journal of Manufacturing Science and Technology 3.2 (2010): 142-157.

[27] Rehman, Sumaira, and Marin D. Guenov. "A methodology for modelling manufacturing costs at conceptual design." Computers & industrial engineering35.3 (1998): 623-626.

[28] Ficko, M., et al. "Prediction of total manufacturing costs for stamping tool on the basis of CAD-model of finished product." Journal of Materials Processing Technology 164 (2005): 1327-1335.

[29] Sol, Henk G., A. Th Cees, and Pieter F. de Vries Robbé, eds. Expert systems and artificial intelligence in decision support systems. Springer, 1987.

[30] Kingsman, Brian G., and Antonio Artur de Souza. "A knowledge-based decision support system for cost estimation and pricing decisions in versatile manufacturing companies." International Journal of Production Economics 53.2 (1997): 119-139.

[31] Partridge, Derek, and Khateeb M. Hussain. "Knowledge Based Information Systems." Mcgraw-Hill Book Comp., 1995.

[32] Gayretli, A., and H. S. Abdalla. "An object-oriented constraints-based system for concurrent product development." Robotics and Computer-Integrated Manufacturing 15.2 (1999): 133-144.

[33] Nikolopoulos, Chris. Expert systems: introduction to first and second generation and hybrid knowledge based systems. Marcel Dekker, Inc., 1997.

[34] Smith, Alice E., and Anthony K. Mason. "Cost estimation predictive modeling: Regression versus neural network." The Engineering Economist 42.2 (1997): 137-161.

[35] Martin, Patrick, J-Y. Dantan, and Ali Siadat. "Cost estimation and conceptual process planning." Digital Enterprise Technology. Springer US, 2007. 243-250.

[36] Asking, Pontus, and Stefan Gustavsson. "Cost breakdown analysis-A study of the product costs in the kitchen appliances at IKEA of Sweden." (2011).

[37] Asking, Pontus, and Stefan Gustavsson. "Cost breakdown analysis-A study of the product costs in the kitchen appliances at IKEA of Sweden." (2011).

[38] Shah, Jami J., and Mary T. Rogers. "Functional requirements and conceptual design of the feature-based modelling system." Computer-Aided Engineering Journal 5.1 (1988): 9-15.

[39] Emmerik, Van, M.J.G.M. and Jansen, F.W. “User interface for feature modelling, in Proceedings” of CAPE '89, Tokyo, October 1989 (F.K. KIMURA & A. ROLSATDAS (Eds), pp. 625-632).

[40] Wierda, Leo S. "Design-oriented cost information: the need and the possibilities." Journal of Engineering Design 1.2 (1990): 147-167.

[41] Wierda, Leo S. "Linking design, process planning and cost information by feature-based modelling." Journal of Engineering Design 2.1 (1991): 3-19.

[42] Zhang, Y. F., Jerry YH Fuh, and W. T. Chan. "Feature-based cost estimation for packaging products using neural networks." Computers in Industry 32.1 (1996): 95-113.

[43] Ou-Yang, C., and T. S. Lin. "Developing an integrated framework for feature-based early manufacturing cost estimation." The international journal of advanced manufacturing technology 13.9 (1997): 618-629.

[44] Leibl, Peter. "Cost calculation with a feature-based CAD system using modules for calculation, comparison and forecast." Journal of Engineering Design 10.1 (1999): 93-102.

[45] Brimson, J. A. "Feature costing: beyond ABC." Journal of Cost Management 12 (1998): 6-13.

[46] Masmoudi, Faouzi, Wafik Hachicha, and Zoubeir Bouaziz. "A new feature-concept applied in cost estimation model for a weld assemblage: Additional Information." (2007): 149-162.

[47] Masmoudi, Faouzi, Zobeir Bouaziz, and Wafik Hachicha. "Computer-aided cost estimation of weld operations." The International Journal of Advanced Manufacturing Technology 33.3-4 (2007): 298-307.

[48] Eklin, Mark, Yohanan Arzi, and Avraham Shtub. "Model for cost estimation in a finite-capacity stochastic environment based on shop floor optimization combined with simulation." European journal of operational research 194.1 (2009): 294-306.

[49] Feldman, P., and A. Shtub. "Model for cost estimation in a finite-capacity environment." International journal of production research 44.2 (2006): 305-327.

[50] Wei, Jingxing, and Y-S. Ma. "Design of a feature-based order acceptance and scheduling module in an ERP system." Computers in Industry 65.1 (2014): 64-78.

[51] Khalifelu, Zeynab Abbasi, and Farhad Soleimanian Gharehchopogh. "Comparison and evaluation of data mining techniques with algorithmic models in software cost estimation." Procedia Technology 1 (2012): 65-71.

[52] Lum, Karen T., Daniel R. Baker, and Jairus M. Hihn. "The effects of data mining techniques on software cost estimation." Engineering Management Conference, 2008. IEMC Europe 2008. IEEE International. IEEE, 2008.

[53] Jørgensen, Magne. "A critique of how we measure and interpret the accuracy of software development effort estimation." First International Workshop on Software Productivity Analysis and Cost Estimation. Information Processing Society of Japan, Nagoya. 2007.

[54] Nguyen, Vu, Bert Steece, and Barry Boehm. "A constrained regression technique for COCOMO calibration." Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement. ACM, 2008.

[55] Hughes, Robert T. "Expert judgement as an estimating method." Information and software technology 38.2 (1996): 67-75.

[56] Gordon, Leonard. "Using Classification and Regression Trees (CART) in SAS® Enterprise Miner TM For Applications in Public Health." Data Mining and Text Analytics (2013): 089-2013

[57] Bidarra, R. A. F. A. E. L., and Willem F. Bronsvoort. "Semantic feature modelling." Computer-Aided Design 32, no. 3 (2000): 201-225.

[58] Ma, Yongsheng, N. Sajadfar, L. Campos Triana "A feature-based semantic model for automatic product cost estimation" IACSIT International Journal of Engineering and Technology, 6(2), 2014, 109-113.

[59] Ma, Yongsheng, T. Tong "Associative feature modelling for concurrent engineering integration" Computers in Industry 51, no. 1 (2003): 51-71.

[60] Al Khalidi, Nesreen, Ahmad A. Saifan, and Izzat M. Alsmadi. "Selecting a standard set of attributes for cost estimation of software projects." In Computer, Information and Telecommunication Systems (CITS), 2012 International Conference on, pp. 1-5. IEEE, 2012.

[61] Chan, Felix TS, and Jie Zhang. "Modelling for agile manufacturing systems." International Journal of Production Research 39, no. 11 (2001): 2313-2332.

[62] Svendsen, Gunnar Lind Haase, and Jens Fyhn Lykke Sørensen. "There's more to the picture than meets the eye: Measuring tangible and intangible capital in two marginal communities in rural Denmark." Journal of Rural Studies 23, no. 4 (2007): 453-471.

[63] Han, JungHyun, Mike Pratt, and William C. Regli. "Manufacturing feature recognition from solid models: a status report." Robotics and Automation, IEEE Transactions on 16, no. 6 (2000): 782-796.

[64] Lin, Than, Jae-Woo Lee, and E. L. J. Bohez. "New integrated model to estimate the manufacturing cost and production system performance at the conceptual design stage of helicopter blade assembly." International Journal of Production Research 50, no. 24 (2012): 7210-7228.

[65] Ulrich, Karl T., and Steven D. Eppinger. Product design and development. Vol. 384. New York: McGraw-Hill, 1995.

[66] Liu, Zhenyu, Wanghui Bu, and Jianrong Tan. "Motion navigation for arc welding robots based on feature mapping in a simulation environment." Robotics and Computer-Integrated Manufacturing 26, no. 2 (2010): 137-144.

[67] Zhang, De-Long, Jiaqing Qiao, Jun-Bao Li, Li-Yan Qiao, Shu-Chuan Chu, and John F. Roddick. "Optimizing Matrix Mapping with Data Dependent Kernel for Image Classification." (2014).

[68] Bernstein, Abraham, Foster Provost, and Shawndra Hill. "Toward intelligent assistance for a data mining process: An ontology-based approach for cost-sensitive classification." Knowledge and Data Engineering, IEEE Transactions on 17, no. 4 (2005): 503-518.

[69] Shaw, N. K., M. Susan Bloor, and A. De Pennington. "Product data models." Research in Engineering Design 1, no. 1 (1989): 43-50.

[70] Peckham, Joan, and Fred Maryanski. "Semantic data models." ACM Computing Surveys (CSUR) 20, no. 3 (1988): 153-189.

[71] Zave, José Nuno Oliveira Pamela. "FME 2001: Formal Methods for Increasing Software Productivity."

[72] Shaul, Levi, and Doron Tauber. "CSFs along ERP life-cycle in SMEs: a field study." Industrial Management & Data Systems 112.3 (2012): 360-384.

[73] Li, Conghua. "ERP packages: what's next?." Information Systems Management16 (1999): 31-35.

[74] Veeramani, Dharmaraj, and Pawan Joshi. "Methodologies for rapid and effective response to requests for quotation (RFQs)." IIE transactions 29.10 (1997): 825-838.

[75] “Epicor Enterprise Resource Planning Overview”,

http://www.atc.gr/files/Brochures/Epicor-Enterprise-Resource-Planning-Overview-ENS.PDF

[76] http://www.infor.com/product_summary/erp/visual/

[77] Sajadfar, Narges, Luis Campos Triana and Ma Y.-S. "Interdisciplinary semantic interactions within a unified feature model for product cost estimation." International Journal of Mechanical Engineering and Mechatronics (2014): 10-19

[78] Cao, Lian Min, Xi Wei Wang, and Qing Liang Zeng. "Study on Cost Modeling Based on Feature Mapping Technology." Advanced Materials Research 295 (2011): 830-833.

[79] Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter 11.1 (2009): 10-18.

[80] Gunduz, Murat, Latif Onur Ugur, and Erhan Ozturk. "Parametric cost estimation system for light rail transit and metro trackworks." Expert Systems with Applications 38.3 (2011): 2873-2877.

[81] Seo, Kwang-Kyu. "A methodology for estimating the product life cycle cost using a hybrid GA and ANN model." Artificial Neural Networks–ICANN 2006. Springer Berlin Heidelberg (2006): 386-395.

[82] Gharehchopogh, Farhad. "Neural networks application in software cost estimation: A case study"

International Symposium on Innovations in Intelligent Systems and Applications (INISTA 2011): 69-73.

[83] Wu, Xindong, and Vipin Kumar, eds. "The top ten algorithms in data mining". CRC Press, 2009

[84] Abbas, Osama Abu. "Comparisons Between Data Clustering Algorithms."International Arab Journal of Information Technology (IAJIT) 5.3 (2008).

[85] Prabhakar, Shashikanth, and Mark R. Henderson. "Automatic form-feature recognition using neural-network-based techniques on boundary representations of solid models." Computer-Aided Design 24.7 (1992): 381-393.

[86] McCoy Drilling and Completions (Farr) Company, product drawings, internal document, 2013.